The Journal of

Applied Psychology

Vol. III DECEMBER, 1919 No. 4

EDITORS

G. STANLEY HALL L. R. GEISSLER

And a Board of Co-operating Editors

WORCESTER, MASS.

FLORENCE CHANDLER, Publisher

Entered as second class matter June 9, 1917, at the post office at Worcester, Mass. under the Act of March 3, 1879.

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. BINGHAM, Carnegie Institute of Technology; JUNE E. DOWNEY, University of Wyoming; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston; H. L. HOLLINGWORTH, Columbia University; JOSEPH JASTROW, University of Wisconsin; HERBERT S. LANGFELD, Harvard University; JAMES H. LEUBA, Bryn Mawr College; HENRY C. McCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, University of Wisconsin; EDWARD K. STRONG, JR., Carnegie Institute of Technology, Pittsburgh; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; J. E. WALLACE WALLIN, Psycho-Educational Clinic, St. Louis; G. M. WHIPPLE, University of Michigan; HELEN THOMPSON WOOLLEY, Vocation Bureau, Cincinnati.

The scope of the JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis (c) The influence of general environmental conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

VOLUME III DECEMBER, 1919 NUMBER 4

TABLE OF CONTENTS

W. D. PARTLOW and THOMAS H. HAINES. Mental Rating of Juvenile Dependents and Delinquents in Alabama 291

HARRY DEXTER KITSON. Comparison Between Two Scales for the Estimation of Intelligence 310

EDGAR A. DOLL. The Average Mental Age of Adults 317

RICHARD H. PAYNTER, JR. A Psychological Investigation of the Likelihood of Confusion Between the Words "Coca-Cola" and "Chero-Cola" 329

EDWARD S. ROBINSON. The Analysis of Trade Ability 352

J. CROSBY CHAPMAN and HERBERT A. TOOPS. A Written Trade Test: Multiple Choice Method 358

S. L. PRESSEY. Minor Studies from the Psychological Laboratory of Indiana University:

III. S. L. PRESSEY and RUTH RALSTON. The Relation of the General Intelligence of School Children to the Occupation of Their Fathers 366

IV. S. L. PRESSEY and I. M. SHIVELY. A Practical Information Test for Use with Delinquents and Illiterate Adults .. 374

V. S. L. PRESSEY. The "Efficiency" of a Group Scale of Inteligence in Prognosticating Success and Failure in Junior High School 381

BOOK REVIEWS 386

NOTES 393

INDEX OF SUBJECTS 397

INDEX OF AUTHORS 399

Communications and manuscripts may be sent to any one of the two editors; reviews, summaries, notices, books and off-prints for review should be addressed to Dr. L. R. Geissler, Clark University, Worcester, Mass. Contributors are requested to follow the conventional method of citation employed by this JOURNAL; in every case the year, volume and pages should be specified.

THE JOURNAL OF APPLIED PSYCHOLOGY is published quarterly. The subscription price is $4.00 a year. Single numbers $1.25. Remittances and business communications should be addressed to Florence Chandler, Publisher, Clark University, (950 Main St.), Worcester, Mass.

The Journal of

APPLIED PSYCHOLOGY

EDITED BY

G. STANLEY HALL L. R. GEISSLER

And a Board of Co-operating Editors

VOLUME III

1919

WORCESTER, MASS.
FLORENCE CHANDLER, Publisher

COPYRIGHT, 1919, BY G. STANLEY HALL

THE BRANDOW PRINTING COMPANY
ALBANY, N. Y.

TABLE OF CONTENTS

GRACE ARTHUR and HERBERT WOODROW
An absolute intelligence scale 118-137

W. V. BINGHAM
Army personnel work 1-12

CHESTER S. CARNEY
National conference on Americanization in industries . 269-276

J. CROSBY CHAPMAN
The learning curve in typewriting 252-268

J. CROSBY CHAPMAN and HERBERT A. TOOPS
A written trade test: Multiple choice method . . . 358-365

EDGAR A. DOLL
The average mental age of adults 317-328

JOSEPHINE CURTIS FOSTER
A case of intellectual development despite enforced seclusion 167-171

GEORGE W. FRASIER
A comparative study of the variability of boys and girls. 151-155

THOMAS H. HAINES and W. D. PARTLOW
Mental rating of juvenile dependents and delinquents in Alabama 291-309

V. A C. HENMON
Air service tests of aptitude for flying 103-109

C. M. KELLEY and F. L. WELLS
Briefer studies from the psychological laboratory of McLean Hospital 172-193

TRUMAN L. KELLEY
Principles underlying the classification of men . . . 50-67

HARRY DEXTER KITSON
Comparison between two scales for the estimation of intelligence 310-316

FRANCES LOWELL
A group intelligence scale for primary grades 215-247

W. D. PARTLOW and THOMAS H. HAINES
Mental rating of juvenile dependents and delinquents in Alabama 291-309

RICHARD H. PAYNTER, JR.
A psychological investigation of the likelihood of confusion between the words "coca-cola" and "chero-cola". 329-351

RUDOLF PINTNER
A non-language group intelligence test 199-214

RUDOLF PINTNER and HERBERT A. TOOPS
Educational differences among tradesmen 33-49

S. L. PRESSEY
The "efficiency" of a group scale of intelligence in prognosticating success and failure in junior high school . 381-385

CONTENTS

S. L. Pressey and L. W. Pressey
The practical "efficiency" of a group scale of intelligence. 68-80
"Cross-out" tests 138-150

S. L. Pressey and Ruth Ralston
The relation of the general intelligence of school children to the occupation of their fathers 366-373

S. L. Pressey and I. M. Shively
A practical information test for use with delinquents and illiterate adults 374-380

S. L. Pressey and G. F. Teter
A comparison of colored and white children by means of a group scale of intelligence 277-282

S. L. Pressey and J. B. Thomas
A study of country children in (1) a good and (2) a poor farming district by means of a group scale of intelligence 283-286

Ruth Ralston and S. L. Pressey
The relation of the general intelligence of school children to the occupation of their fathers 366-373

Edward S. Robinson
The analysis of trade ability 352-357

I. M. Shively and S. L. Pressey
A practical information test for use with delinquents and illiterate adults 374-380

Walter B. Swift
Mental attitudes in debate 156-166

G. F. Teter and S. L. Pressey
A comparison of colored and white children by means of a group scale of intelligence 277-282

J. B. Thomas and S. L. Pressey
A study of country children in (1) a good and (2) a poor farming district by means of a group scale of intelligence 283-286

Edward L. Thorndike
A standardized group examination of intelligence independent of language 13-32

L. L. Thurstone
Mental tests for prospective telegraphers 110-117
A standardized test for office clerks 248-251

Herbert A. Toops and J. Crosby Chapman
A written trade test: Multiple choice method 358-365

Herbert A. Toops and Rudolph Pintner
Educational differences among tradesmen 33-49

F. L. Wells and C. M. Kelley
Briefer studies from the psychological laboratory of McLean Hospital 172-193

Herbert Woodrow and Grace Arthur
An absolute intelligence scale 118-137

In Memoriam—John Wallace Baird . . . . . Frontispiece

A Symposium. The field of clinical psychology as an applied science 81-95

Notes 195-198, 289-290, 393-396

Book Reviews 96-102, 194, 287-288, 386-392

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. IV DECEMBER, 1919 No. 4

MENTAL RATING OF JUVENILE DEPENDENTS AND DELINQUENTS IN ALABAMA.

By W. D. PARTLOW, M. D., Superintendent of Alabama Insane Hospitals, Tuscaloosa, Alabama; and

THOMAS H. HAINES, M. D., Professor of Medicine (Nervous and Mental Diseases), College of Medicine, Ohio State University, Columbus, Ohio.

This paper constitutes a report upon the value of group intelligence rating tests, and the relations of the same to individual examinations made by the Binet-Simon method as a part of a mental survey of the four industrial schools of Alabama. This opportunity came to us by request of the Chairman [1] of the newly constituted Board of Control and Economy. This Board recognizes it to be a matter of fundamental economy for it to know the kinds of mental equipment the boys and girls of these industrial schools possess, in order that the real work of the institutions may be forwarded in the best possible manner.

The schools investigated are four in number,—the Alabama Boys' Industrial School, at East Lake, with a population of about 307 white boys, the State Training School for Girls and the Mercy Home Industrial School for Girls, both at Birmingham, with about 45 and 30 white girls respectively, and the Alabama Reform School for Juvenile Negro Law-Breakers, at Mt. Meigs, with 264 colored boys. These children range in age from seven years to twenty.

[1] The request for this work on the part of the State Board of Control and Economy was met by the Alabama State Board of Health and the National Committee for Mental Hygiene. These agencies financed the survey.

Backwardness of Industrial School Population

It is a well known fact that boys and girls selected by Juvenile Courts, and other social agencies, for residence at such institutions as these industrial schools, are backward in school work,—that they are in large measure what we call dullards and feeble-minded.

TABLE I

Distribution by age and School Grade at the Alabama Boys' Industrial School

School Grades	Ages in Years													Totals
	7	8	9	10	11	12	13	14	15	16	17	18	19	
Primer	1	1	2		1						1			6
I		4	2	2	2	2	4	2			2	1	1	22
II			3	2	7	6	6	7	4	4	9	1	2	51
III			1	4	6	6	7	14	8	4	2	2		54
IV		1			4	3	15	12	6	7	3			51
V				1	1	3	5	13	12	10	1	2		48
VI							3	10	9	7	9	1		39
VII								1	8	3	3	1		16
VIII							1	2	6	2	2			13
IX									1	1	2			4
Totals	1	6	8	9	21	20	41	61	54	38	34	8	3	304

Table I exhibits these facts from the point of view of school classification for the Alabama Boys' Industrial School. 304 of the 307 boys examined, for whom school classification was stated, are classed according to age and school grade. If it is fair to assume that a child in his seventh year should be in the primary grade, in his eighth year in the first grade, in his ninth year in the second grade, and in his fourteenth year in the ninth grade, we find that there are only eight of these

304 children who are up to, or beyond, the expected grade. They are shown below the diagonal line in the table. Only 30 other children of the 304 are in the grades immediately below these, or one year retarded, according to these expectations. The great mass of children point in the direction of the nine second-grade seventeen year-olds.

The modal age is fourteen years, and the modal grade is the third. Likewise the maximal number of third graders is fourteen at fourteen years of age. The modal grade for the fourteen year-olds is the third, and the modal age in the second grade is seventeen years with nine already alluded to. Likewise the median age of the second grade is fourteen years.

The group of this Industrial School population is, in fact, so loaded with retarded children that this condition cannot be attributed to the accident of poor school facilities. Especially true is this when we consider what large numbers of these boys come from the cities where excellent school facilities are provided.

Group Intelligence Ratings

Experience with group tests in the applications of psychology to the organization of military forces had convinced us of the labor-saving feature of this device in such a survey as this. The military tests were not available when this work was planned. Considering the large amount of illiteracy amongst these boys and girls, as indicated by the large numbers classed in the lower school grades in Table I for the white boys, it seemed advisable to prepare two forms of group examination, one for children of the fourth school grade degree of literacy in English, and the other for children of the third school grade and below. The form prepared for those able to read English, and designated 'Form A' of Group Intelligence Rating Tests, is constructed very much on the lines of the Alpha Examination used in the Army. It is somewhat shortened and introduces one performance test, namely, the mazes. It consists of eight parts, (1) Directions Test, (2) Arithmetical Performance, (3) Mazes, (4) Same-Opposite, (5) True-False, (Disarranged Sentences), (6) Analogies, (7) Choice of Reasons for Stated Facts, and (8) Arithmetical Series. The total possible score for these eight tests is, according to the method adopted, 193, and the actual working

time of the examinee about twenty minutes. It is, therefore, an easy matter to go through the whole performance with a group, including the filling in of the headings of the paper, in the space of an ordinary school period of forty-five minutes.

The other form of Group Intelligence Rating, Form X, designed for children who have not much reading knowledge of English, consists of five parts, the first four of which involve no reading of English whatever. The first is a Directions Test, the second is a game at store keeping, marking goods in letters which stand for figures (a letter digit test), the third is a maze test, and the fourth consists of twenty unfinished pictures. A cloth blackboard is carried to aid in the explanation of the second, third, and fourth tests. The first maze is on the blackboard entire, and a child is asked to volunteer to show the group how to get through this maze, by coming up and taking the chalk and making a mark from the arrow on the left to that on the right of the maze. The first unfinished picture, a girl without a mouth, is on the blackboard as on the examination book, and a volunteer is asked to come up and finish this picture. The transition from unfinished pictures to unfinished sentences (Test 5) is made as easy as possible by the fact that both are unfinished. The unfinished sentence is by all odds the most difficult of these five tests for these relatively unlettered children, because it involves reading of English. The total possible score in this Form X of Group Intelligence Rating is 182. The actual working time of the examinee is about seventeen minutes.

In order to be as economical of our time as possible, and at the same time disturb the routine of the institutions as little as possible by our work, we proceeded first to make group examinations of the intelligence of the boys and girls of these schools by using these two forms.

The same two forms were used about the same time on an aggregate of 1012 children in the schools of Jackson, Mississippi, ranging in age from five to thirteen, and in school grade from the primer grade to the seventh, or highest grammar school grade. In the same way 765 colored children in the schools of Jackson, ranging in age from five to fourteen, and in school grade from the primer to the seventh, were rated as to intelligence by these two forms. The percentile distribution of the scores made by these Jackson school children constitute our measuring instrument for the performances of the boys and girls, white and colored, of the Alabama Industrial Schools.

Group Intelligence Ratings of Children in Alabama Industrial Schools

TABLE II

Distribution of Scores in A & X Forms of Group Intelligence Rating by Age of Child and by Percentile Maxima of Public School Children. Alabama Boys' Industrial School

Percentile	Ages in Years													Totals
	7	8	9	10	11	12	13	14	15	16	17	18	19	
100					1		1							2
90							1	1	3	2	4	1		12
80				1				2	4	4				11
70							1		4	2	2			9
60							2	2	2	2	2			10
50					1	1	2	3	3	1	2			13
40			1		1	1	2		4		2			11
30				1		1	1		1	2	2			8
20		1	1		2	3	2	10	8	7	2			36
10			1	2	4	5	2	17	4	5	2	2		44
5				2	5	2	4	8	6	3	4	1		35
3		4	3	1	1	4	2	2	1	3	1			22
2	1	1	2	2	6	4	22	16	15	7	11	4	3	94
Totals	1	6	8	9	21	21	42	61	55	38	34	8	3	307

Table II presents the percentile distribution of the performances of 307 boys of the Alabama Boys' Industrial School rated by the percentile distributions of performances of Jackson school children. Since we had very small numbers of children beyond nine years of age in the third grade in the public schools, nine years is the highest for which we have reliable percentiles in the X form of examination. Likewise our figures for the A form of examination comprise children from nine to thirteen years, inclusive.

219 of these 307 boys are beyond thirteen years of age. More than forty per cent of these boys took the X form of

examination by virtue of their retardation in school work. There is necessary, therefore, a certain amount of projection of expectation of performance with advancing age. We have to anticipate what a thirteen year-old might reasonably be expected to do in the X form of examination from what seven, eight, and nine year-olds have done, and likewise we have to anticipate what a sixteen year-old might be expected to do in both the A and X form from what younger children have done.

The distribution of the scores of the Boys' Industrial School inmates have been made in Table II according to this simple plan. Any score falling below the lowest two percentile performance of Jackson school children is placed in the two percentile group under the age of the child making it. Any score falling above the two percentile, and below the three percentile limit, is placed on that level, and so on, above three and below five is placed in the five level, and above five and below ten is placed in the ten level. By this simple plan of stratification we have put before us another striking exhibition of the retardation of the minds of these boys. 151 of these boys put up mental performances below the limits of the lowest five per cent of public school children of the same ages and race. That forty-nine per cent of these boys are so incapable mentally is really a more striking exhibition of the problems confronted by the management of such an institution than are the actual numbers of feeble-minded which we present later.

This classification of intelligence of the boys abundantly justifies the school grading as set forth in Table I.

Of course we have constantly to bear in mind the fact that the group examination does not detect the lazy individual who will not put forth his best effort under the conditions of such examinations. However, the conditions of the test were exactly the same in the public schools and in these industrial schools.

In the same way Table III exhibits the distribution of scores of the 75 girls of the two Training Schools according to the percentile stratification of performances of public school children.

33 of the 75 girls fall below the five percentile level of public school children, and only one of the whole number makes a score above the median performances of public school children, whereas 44 of the 307 boys exhibit scores above the median performances of public school children. This seems to indicate in general a lower grade mentality amongst the white girls of these reform schools than amongst the white boys.

TABLE III

DISTRIBUTION OF SCORES IN X FORM OF GROUP INTELLIGENCE RATING BY AGE OF CHILD AND BY PERCENTILE MAXIMA OF SCHOOL CHILDREN. STATE TRAINING SCHOOL FOR GIRLS AND MERCY HOME INDUSTRIAL SCHOOL FOR GIRLS

Per-centiles	Ages in Years														Totals
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
100															
90															
80															
70															
60							1								1
50						2	1	1	2		3	1			10
40						1			1	2					4
30							2	1	1	1	1	1			7
20							1	2	4	4	1	2			14
10							1		2	2	1				6
5						2	1	1	2	1	1				8
3															0
2	1		1		1		3	2	6	5		2	3	1	25
Totals	1		1		1	5	10	7	18	15	7	6	3	1	75

Table IV shows the stratification of the scores of 264 negro boys in the Alabama Reform School for Juvenile Negro Law Breakers. This stratification is made by comparison with the performances of 765 negro children in the public schools of Jackson, Mississippi. It is a very different distribution from what would be presented had they been compared with the performances of white children. 221 of these 264 boys, or 84%, score below the median performances of children of their same ages and race. Judged by the standards of their own race, 84% of these boys are below the average of intelligence. 16% therefore, are above average intelligence for their race, whereas 14% of the white boys, and 1% of the white

TABLE IV

DISTRIBUTION OF SCORES IN A AND X FORMS OF GROUP INTELLIGENCE RATING BY AGE OF CHILD AND BY PERCENTILE MAXIMA OF PUBLIC SCHOOL CHILDREN. ALABAMA REFORM SCHOOL FOR JUVENILE NEGRO LAW BREAKERS

Percentiles	Ages in Years												Totals
	8	9	10	11	12	13	14	15	16	17	18	19	
100				2			1		1				4
90				1			1	1	1				4
80			3			2	4		1				10
70					2			5	3	1			11
60			2	1	1	3	2	1	2	1	1		14
50		1				5	4	2	1				13
40				5	2	3	5	4	2	1			22
30		1		1	8	15		3	2	1			31
20	1			1	7	16	11	15	14	4		1	70
10			2		3	3	15	10	12	10	2	1	58
5			1	2	3	1	7	5	5	1	1	1	27
Totals	1	2	8	13	26	48	50	46	44	19	4	3	264

girls are above the average intelligence for their ages and race.

85 of the colored boys, or 32% of the total number score below the limits of the lowest 10% of public school children of their race, and 27, or 10% below the limits of the lowest 5%.

The scores of the colored boys do not bunch so closely in the lowest five percentile strata as do the white boys. A much larger number is found between the ten and twenty percentile limits than below the five percentile limit.

But it is evident that all three of these groups,—white boys, white girls, and colored boys, are over-loaded with dull children, white boys 86% below average, white girls 99% below average, and colored boys 84% below average.

The distribution, however, within this lower half of the field of mental ability is very different for the three groups.

Below the ten percentile level we find white boys 64%, white girls 52%, and colored boys 32%.

Below the five percentile level we find white boys 49%, white girls 44%, and colored boys 10%.

On being compared with school children of the same race, the colored boys are, therefore, from all three of these angles, better endowed than the white children. While they have not significantly larger numbers, or percentages, above the average intelligence of their race, they do rank higher in the low groups, or strata. Whatever is the selection process of the courts sending these boys and girls to these schools, they evidently do not select negro children so closely as they do white children along the lines of racial dullness, or lack of intelligence.

Individual Examinations

Throughout the experience of the personnel of the Division of Psychology in the Surgeon General's office in separating the feeble-minded from the competent in the draft army, it was found necessary to follow the scoring of group performances by individual intelligence examinations in order to make satisfactory diagnoses and to accumulate the evidence necessary for discharge from the service on account of mental inferiority. For the most part the feeble-minded were found in the so-called 'E Group' of scores resulting from group intelligence rating. This group constituted roughly the lowest five per cent of the draft.

As to percentages found Terman states that 16,266 men out of 1,059,767 examined between May 1, 1918 and September 1, 1918, had a mental age of eight years, or less.[2] This constitutes one and six-tenths per cent of the gross number in the drafts under consideration.

Yerkes[3] states later that of 1,726,000 men examined prior to January 1, 1919, 45,563 were of a mental age of less than ten years. This is 2.6% of the drafts under consideration. As, however, only 27,052 of these were recommended for (1) discharge as feeble-minded, (2) labor battalions, or (3) development battalions, this is probably the truer measure of mental deficiency as found in this draft. This number, 27,052, is 1.6% of the drafts under consideration. It seems likely, therefore, that something like one and one-half per

[2] L. M. Terman, *Psychological Bulletin* XV, June, 1918, pp. 177-187 "The Use of Intelligence Tests in the Army."

[3] R. M. Yerkes, *Science* March 7, 1919, p. 226, "The Measure of Utilization of Brain Power in the Army."

cent of the men who passed local exemption boards and were conducted to camps as reasonably suitable material for making fighting men were so defective in mentality that there was no hope of developing them into serviceable factors in the military organization.

The defects of the plan of rating adults by mental age, and judging of their efficiency thereby, are exhibited by these figures. It is evident, despite the striking presentation it enables one to make to the lay mind of the defect of a mentally inferior person, that it does not represent the facts in the case to say of a given individual twenty-five years old that he has the mind of a child of seven, or of nine years.

The different ages represented in the developmental period by these children in the Alabama Industrial Schools complicates our problem considerably beyond that confronting the military psychologist. We have to treat each age group up to sixteen as a distinct group and to compare the score of each with the scores of normal children of the same age and race.

As a matter of practical procedure, being compelled to proceed with our individual examination work before the stratification of scores of public school children was completed, we arranged the scores in each age group, for each race and sex, and for each form of the examination, in order of their magnitude, and proceeded with the children making the lowest scores up in the lists until we felt satisfied we had reached a point beyond the level of feeble-minded performance. This empirical procedure probably involved a little more labor than would have been necessary had we had at our disposition the stratification of public school children's performances. However, with this classification of public school children's performances in hand we would have proceeded in exactly the same way, only being enabled to stop a little short of the point to which we mounted in each group. For mental age rating we used the Stanford Revision of Binet-Simon Tests.

Value of Group Tests

The first point in favor of the use of group intelligence rating methods in such institution surveys as these herein reported, consists in the facilities afforded for a quick and fairly accurate intelligence rating of every individual in the survey. It is a matter of decided value to the Superintendent and management of an institution, just as to the personnel officer in a division of the Army, to know of each individual that he is above, or below the average in intelligence, and how much above or below he is. It is of service to such management

to know that one-half of the children are as low down in ability to adapt themselves to their environment and to learn how to do things as the lowest five per cent of public school children of their same ages and race.

There is no question about the labor saving effected by this device when one is setting about the survey of an institution population with the specific object in mind of sorting out the feeble-minded. It is a much more effective procedure to make group tests and get the scores than to discuss intelligence capacity and possibility of mental deficiency with the Superintendent and teachers. This situation has not changed in the twenty years since Binet called attention to the very poor ability of the average teacher to diagnose, or to aid in the diagnosis, of mental deficiency in the case of any child in her room. Any who have ever tried this method of discovering which children in a given group should be submitted to an individual examination for mental age rating and social prognosis know how exceedingly unsatisfactory it is. Such persons know that one must practically examine every child in order to be reasonably certain he has not passed over some mental defectives.

By this method we found it unnecessary to extend our individual examinations beyond about one-third of the white boys and white girls, and a little over one-half of the colored boys. This is some curtailment of labor.

Of the boys and girls actually examined individually, approximately one half were found to be feeble-minded.

As stratification of intelligence of school children proceeds to larger numbers, and we can state with more definiteness the expected performance of the lowest two per cent, the lowest three per cent, and the lowest five per cent of children of each age through the developmental period, we shall be able to come closer to the feeble-minded group by the group method of intelligence assessment. As this approximation is made, more labor in individual examination will be saved.

Of course this is an asymtotic approximation, the special factor precluding diagnosis of mental defect by the group method alone being the impossibility, already referred to, of detecting that species of volitional interference with the production of good results in an individual, otherwise capable, under the conditions of the group method of intelligence testing. There will always be individuals who will do less than they are capable of performing. Some such will, in any group, have to be interviewed and examined individually in order to demonstrate that they are not as poorly equipped as their group intelligence rating results indicate.

TABLE V

RELATION OF GROUP SCORES TO DIAGNOSES RESULTING FROM INDIVIDUAL EXAMINATIONS

TOTAL NUMBERS OF

(1) Scores as of lowest five percentile public school children.
(2) Children assigned mental age by Stanford Binet-Simon Test.
(3) Diagnosed as Feeble-minded (institution cases).
(4) Diagnosed as Border-line (probably institution cases).
(5) Diagnosed as not Feeble-minded as result of individual examination and mental age rating.

	White Boys	White Girls	Colored Boys
Numbers scoring in Group Rating as of Lowest 5 Percentile	151 (49%)	33 (44%)	27 (10%)
Numbers Assigned Mental Age by Stanford Binet-Simon Examintion	94 (31%)	25 (33%)	140 (53%)
Numbers diagnosed as Feeble-minded	44 (14%)	16 (21%)	57 (22%)
Numbers diagnosed as of Border-line Intelligence	13 (4%)	3 (4%)	42 (16%)
Numbers diagnosed as not Feeble-minded by Individual Examination.	37 (12%)	6 (8%)	41 (16%)

Table V exhibits for the three classes of children examined the total numbers rating themselves below the five percentile limit of public school children of the same race, and, in parentheses beside these figures the percentages which these numbers constitute of the total numbers of those examined. The next line presents the numbers and percentages of each group submitted to individual examination; the next, the numbers and percentages of each group diagnosed as feeble-minded; the next, the numbers and percentages diagnosed as of border-line intelligence; and the last, the numbers and percentages submitted to individual examination but pronounced not feeble-minded. It should be stated parenthetically at this point that considerable numbers of boys and girls

submitted to individual examination and found not to be feeble-minded were found to be constitutionally inferior persons, or epileptics, or psychopaths, and for these reasons likely to prove constant public charges on account of criminal tendencies growing out of their abnormal characters. Many children were seen individually and partial mental examinations made, for whom no mental age assignments were made.

An out-standing feature of this Table is the very large percentage of negro boys given individual examinations. It will be noticed also that a considerably larger percentage of these negro boys are diagnosed as feeble-minded than of the white boys, and a very much larger percentage of the negro boys are diagnosed as of border-line intelligence than of either white boys or white girls. It is a fact that we are without data in regard to the reasonably expected performances by a Binet-Simon form of examination of negro children at the various periods of development. The standardization of individual intelligence rating has been done with white children. We do not know what to expect of negro children. As a consequence of this defect in our methods of applied psychology we found ourselves more at a loss in making prognoses of these negro boys in Alabama than in respect to the white boys and girls.

The group tests with Jackson school children, and the work with group tests in the Army, have shown unquestionably the backwardness and inferiority of the mental development in the negro as compared with that in our white population, but we are entirely without a guide as to the measurement of this inferiority in terms of years and months. We have no standards by which to judge what degree of intelligence is required of a negro boy of a given age in order to warrant us in predicting his ability to be such that he will be able to manage himself and his affairs with prudence in this civilization of ours. Consequently, we have taken refuge in classing a larger number of them as of border-line intelligence. At the same time we have found it necessary to examine individually a far larger number of them, and have classed a larger percentage of them as feeble-minded than of either white boys or white girls.

Certain it is that if we judge these negro boys by the standards established for white children, we should find two thirds of the population of the Alabama Reform School for Juvenile Negro Law Breakers subjects for state care as mental defectives, and we should also find very large percentages of negro children in the public schools subjects for state care.

Manifestly the practical results of allowing negroes of such inferior intelligence, judged by white standards, to undertake to manage for themselves, are not disastrous. We are not likely to have to look after larger percentages of negroes as state wards than we do of the white population. It is, therefore, a scientific desideratum of the highest import that we should have a standardization of the Binet-Simon Tests to the development of negro intelligence.

Another feature of contrast between the races manifest in our results is a wider percentile distribution of the negroes as distributed according to the strata of performance of public school children. Of the white boys and girls given individual examinations, only three had scores in the ten percentile group, that is, above the five percentile limit, two of them boys, eleven and twelve years old, made, on individual examination, intelligence quotients of 100 and 104; the other, a girl seventeen years old, made an X score of 102, but on individual examination made an intelligence quotient of 59, her mental age being 9.4. This girl, it should be said, was called in for individual examination more on the strength of a statement by the Superintendent than on account of her showing in the group tests, and she was rated as a psychopathic personality really suffering from an incipient psychosis. As a result of this condition she shows two different mentalities, one in the group test which is good, and another in the individual examination which is inferior.

On the other hand, the negro boys called for individual examination and found to be feeble-minded range high in the percentile stratification of their group scores.

Table VI exhibits these facts, showing that the majority of the five percentile group (63%) were diagnosed feeble-minded, only 43% of the ten percentile group, 20% of the twenty percentile group, and 3% of the thirty percentile group. Likewise, decreasingly small proportions of these groups were called for individual examination.

The distribution of examinees and the resulting diagnoses of these negro boys shown in Table VI indicate that the standard of feeble-mindedness for the negro boy, which the examiners had in mind in their work, cuts through the percentile distribution of group scores of negro children at about the twenty percentile level. This twenty percentile level in negro children seems to correspond closely to the two percentile level in white children. 28% of the five percentile white boys are found feeble-minded, and none above the five percentile group. 45% of the five percentile white girls are found feeble-

minded, and, as just explained, no real case is found above this group. On the other hand 36% of the lowest twenty percentile group of the negro boys are feeble-minded, and one case is found above the twenty percentile level.

TABLE VI

DISTRIBUTION OF GROUP SCORES OF NEGRO BOYS GIVEN INDIVIDUAL EXAMINATIONS, AND NUMBERS OF EACH GROUP FOUND FEEBLE-MINDED, BORDER-LINE AND NOT FEEBLE-MINDED

Percentile Group	Numbers in Groups (Table IV)	Numbers Examined Individually	Found Feeble-minded		Diagnosed Border-line	Diagnosed Not Feeble-minded
			Numbers	Per cents of whole groups		
5 Percentile	27	27	17	63%	8	2
10 Percentile	58	51	25	43%	14	12
20 Percentile	70	47	14	20%	17	16
30 Percentile	31	13	1	3%	2	10
40 Percentile	22	1			1	
80 Percentile	10	1				1

REASONS FOR THIS SCATTERING OF THE NEGRO FEEBLE-MINDED

We cannot satisfactorily explain this disparity of results of correlation between group and individual examinations for the white and colored populations. The following facts, however, may be taken in partial explanation of this disparity:

(1) Smaller numbers of negro children were examined in the public schools; 765 stratified scores as contrasted with 1012 for white children. It was also noted in this stratification that there were numerous cases of maxima for older children falling below maxima for younger children at the same levels.

(2) It is a question how far the streaks of brilliance found

occasionally in mulatto children may interfere with the work of standardization in such a school population. Of course large numbers will overcome this, but it is a distinctly different element from any found in a white school population.

(3) Unquestionably the greatest source of error, and the greatest factor in producing this disparity of results is the *uncertainty of age* of negro children. Both in the public schools and in the reformatory we find negro children absolutely ignorant, not only of date of birth, or time when birthday comes, but also of the numbers of years of age. Constantly such children and their teachers guess at the age. This ignorance and apparent attempt to deceive in regard to age so impressed us at the Juvenile Negro Reformatory that we came to rely almost entirely upon our own judgment of the age of the boy from what we could observe of his size, growth, and physical development, especially of the secondary sex characters. Facts of this sort observed seemed to us very much more reliable than the statements of the boy. No reliable school records were found in either the Reformatory or in the public schools. A special factor making for the unreliability of stated age in the Reformatory consisted in the fact that boys overstated their age with the hope of early dismissal from the institution.

(4) The more tentative character of diagnosis of feeble-mindedness in the negro child as a result of individual mental examination and age rating, which has already been referred to, constitutes another possible source of this disparity stated for the two races.

The Two Per Cent Strata of Intelligence In White Children

Of the 94 white boys given definite mental age rating as a result of individual examination by the Binet-Simon method, 43 were found to be feeble-minded, and 41 of these had made scores by the group tests falling below the lowest two percentile level of public school children. The other two found feeble-minded were between the two and three percentile level of public school children in their group scores.

Of 13 white boys diagnosed as of border-line intelligence, 10 in their group scores fell below the two percentile level of public school children; two others below the three percentile level; and the remaining one below the five percentile level.

The remaining 38 of these 94 white boys who were diagnosed as not feeble-minded are divided as follows in regard to

their percentile ratings. 24 fell below the two percentile level; 9 others below the three percentile; three others below the five percentile level; and two below the ten percentile level. From this it is evident that all the feeble-minded would be caught in the net of the group tests by adhering to the three percentile level. Only six boys who made scores above this level were seen in individual examination. One of these was diagnosed as of border-line intelligence, and the remaining not feeble-minded. It is likewise evident that practically all the feeble-minded would be found by further sorting over by the individual examinational method those boys scoring in the group tests below the two percentile level.

In the two schools for white girls, 25 girls were called for individual examinations and given mental age ratings. 21 of these were below the two percentile level, the remaining 4 being below the five percentile level. None of these girls rated between the two and three percentile level.

15 of these 21 girls below the two percentile level in group tests were found to be feeble-minded; 3 of border-line intelligence; and 3 not feeble-minded. Of the 4 girls rating as of five percentile intelligence, one was found to be of border-line intelligence, and three not feeble-minded, as the result of individual examinations.

In these three groups of white children numbering 382 in all, 119 scored by the A or the X form of group intelligence rating below the limit of the lowest two per cent of 1012 white public school children.

Had the individual examinations been confined to these 119 children, we would have found 56 of the 58 whom we diagnosed as feeble-minded as the result of individual examination, and 13 of the 17 for whom we made provisional diagnosis as being of border-line intelligence.

Using the group tests as the first screening process, and assuming the two percentile limit to have corralled practically all of the feeble-minded, it is seen we would have saved a considerable amount of effort by this use of the group method of intelligence rating. This two per cent would comprise a considerable number of dullards, and some, perhaps, who are not dull, but for one reason or another would not put forth their maximal effort under the conditions of the group test. Many of these, too, are defective characters,—constitutional inferiors, who constitute most serious problems for their caretakers. These need most careful attention from the psychiatrist. Others of this class will be picked from those making

higher scores by their behavior, and by the reported difficulty they cause the management.

Individual mental examination made in this two percentile group, which in these schools constitutes less than 40% of the population, will clear, apparently, about two-fifths of the individuals from suspicion of being feeble-minded. It will pronounce them able to learn and to adapt themselves to their environments so as to be able to manage themselves and their affairs with prudence, and will, therefore, consider them as constituting no danger to themselves, or to others on account of defects of intelligence.

About one half of the two percentile group of these white children prove to be clearly feeble-minded in the sense of this definition just given. The remaining ten per cent of the group we found it necessary to provisionally diagnose of borderline intelligence. We mean by this that our observation has not been sufficient to enable us to pronounce the child so designated as feeble-minded, or as not feeble-minded. We must leave the definite diagnosis to further experience with the child. In many cases we feel that a longer period of experimentation with him is necessary before a definite diagnosis can be given. In fact, with many children of this class the pragmatic test of experience alone will demonstrate whether or not the individual is simply a dullard, defective in some range of imagination, and entirely capable of managing himself and his affairs with prudence, or whether he is so defective through inheritance that he will have to be looked after by the community in order to keep him straight and to save the expense of his crimes and pauperism.

From this point of view the group tests work out very much as would be expected. Army results show about one and one half per cent of the young adult male population to be feeble-minded. It is usually conceded that about one per cent of the total population are feeble-minded in the sense that they have not intelligence enough to manage themselves and their affairs with prudence, and hence need community care to save the expense of their own short-sightedness. If these group tests really measure the intelligence of the children, it should be possible to find all of those who are so defective as to need community care within the group of the lowest two percentile level of public school children. Finer work in the future with group tests may possibly bring us yet nearer to an ultimate sorting out of the feeble-minded by the group method of examination alone.

Without doubt the group intelligence ratings of negro chil-

dren can be made much more serviceable when a means is found of ascertaining definitely the ages of children examined.

Summary

(1) Group Intelligence Rating scores of white boys, white girls, and colored boys, of Alabama Industrial Schools are stratified by the percentile achievements of 1012 white children and 765 colored children of the schools of Jackson, Mississippi.

(2) Only 8 of the 307 white boys are up to the expected school grade. Only 30 others are as little as one year retarded.

(3) The following percentages score below the 50% level of public school children: White boys 86%, white girls 99%, colored boys 84%.

(4) The percentages below the ten percentile level for the three groups in the above order are 64%, 52%, and 32%.

(5) The percentages below the five percentile level for the three groups in the above order are 49%, 44%, and 10%.

(6) The feeble-minded white boys and white girls, and those provisionally diagnosed as of border-line intelligence are practically all found among those scoring below the lowest two per cent of public school children.

(7) This shows the group test to be a labor-saving device in such mental surveys of institutions. Three-fifths of the lowest two percentile scores are feeble-minded or of borderline intelligence in the three schools for white children surveyed.

(8) A larger percentage of negro boys had to be examined individually in order to catch the feeble-minded, and these boys' scores in the group tests range up to the thirty per cent level of intelligence of colored children in the public schools. In fact the twenty per cent level in colored children seems to correspond closely to the two percentile level in white children as a group score limit of feeble-minded intelligence.

(9) Ignorance of actual age in colored children may be responsible for this disparity of results.

(10) Uncertainty of the applicability of Binet-Simon standards to the mind of the negro child may constitute another reason for this disparity in the relations of group and individual intelligence examination results in the two races.

(11) Numbers and percentages diagnosed feeble-minded for the three groups are as follows: White boys 44 (14%), white girls 16 (21%), colored boys 57 (22%).

COMPARISON BETWEEN TWO SCALES FOR THE ESTIMATION OF INTELLIGENCE

By HARRY DEXTER KITSON, Indiana University

In view of the prevailing interest in methods of estimating abilities, the writer undertook a comparative study of two varieties of scales for the estimation of intelligence.

The first, devised by Pearson, has apparently not been used in America except at The University of Chicago where it is a part of the program of psychological investigation which the writer carried on for a number of years.[1] Some data relative to the use of the scale in this connection have been presented by Ruml.[2]

The second scale was first used[3] in connection with psychological investigations upon the ability of salesmen. The idea formed the basis of the Army Officer's Rating Scale and is now being used as the basis of several other rating scales in process of development.

The Pearson scale (hereafter designated, Scale I) is described at length in *Biometrika*,[4] but lest the files of that journal be inacessible to some readers a summary will be given here. The scale consists of a number of classes or degrees of intelligence which were defined by Pearson substantially as shown in the form below. On the basis of this classification a number of headmasters and teachers in English schools were requested to rate their pupils. The persons rated included children in grammar school and students in Oxford and Cambridge, totalling about 5,000.

Assuming that the distribution of intelligence follows the curve of probability, Pearson plotted the estimations along this curve; found the numerical value for each class by determining the point on the scale corresponding to the mean value; determined the limits by finding the deviations from the mean

[1] Kitson, H. D. The Scientific Study of the College Student. *Psychol. Rev. Mon. Sup.* 1917, Vol. 23, No. 1.

[2] Ruml, B. The Reliability of Mental Tests in the Division of an Academic Group. *Psychol. Rev. Mon. Sup.* 1917, Vol. 24, No. 4.

[3] "Aids in Selecting Salesmen," published by Bureau of Personnel Research, Carnegie Institute of Technology, November, 1916.

[4] Pearson, K. Relationship of Intelligence to Size and Shape of Head. *Biometrika*, Vol. V., 1906, pp. 105-46.

(measured in terms of the S. D.) required to include in each class the observed frequency.

The S. D. was approximately 100, and so the range of the two classes marked "Fairly Intelligent" and "Slow Intelligent" was divided into 100 parts called *mentaces*. Pearson called average intelligence Zero, and expressed standings above and below by plus and minus scores. The numerical values of the classes were found by determining the number of cases falling between the averages for each two classes (based on the probability integral). These values are shown in Table I. This table differs from the original one of Pearson's, in placing average intelligence at 300 and indicating deviations above and below, by steps; thus eliminating the use of plus and minus signs to indicate deviation above and below the average. It also contains intermediate steps between the mid-points of Pearson's classes allowing finer discrimination for such judges as care to use them.

TABLE I

Numerical Values of Classes and Sub-Classes			
M+471	O+337	Q+220	S+96
M 451	O 322	Q 192	S 62
M—416	O—302	Q—177	
N+391			
N 371	P+282	R+157	
N—353	P 262	R 130	
	P—242	R—116	

At the close of each quarter of the university year instructors of students in the School of Commerce and Administration are given the blank reproduced herewith, with the request that they rate their students. Each student thus receives three estimations in the course of nine months, thirty-six during his entire course. These are added and averaged, thus giving basis for a rank upon the scale of intelligence.

Classification of Intelligence

In order to standardize judgments on mental ability, the following classification of intelligence has been selected. Please note that the scale covers the range of the population at large from the genius to the imbecile.

On the accompanying blanks, please place the letter standing for the class in which you judge the student's mental capacities to fall. What is desired is a judgment of general intelligence and not of classroom performance. If the in-

dividual seems to be in one class with marked leaning toward another class, judge him to be in the more certain class and indicate the direction by + or —.

Class L—Genius

Class M—Especially Able: A mind especially bright and quick both in perception and in reasoning; able and accustomed to reason rightly about things on pure self-initiative, even when the data are novel.

Class N—Capable: A mind less likely than M to originate inquiry, but quick in perception and reasoning.

Class O—Fairly Intelligent: A mind ready to grasp and capable of perceiving facts in most fields; capable of good reasoning with moderate effort. This group comprises, say, one-third of the total population.

Class P—Slow Intelligent: A mind slow generally, although possibly more rapid in some fields, but quite sure of knowledge once acquired.

Class Q—Slow: A mind very slow, but not incapable of progress, granted time and considerable effort.

Class R—Slow Dull: A mind capable of perceiving relationship between facts in some few fields with long and continuous effort, but not generally or without external aid.

Class S—Very Dull: A mind capable of holding only the simplest facts, and incapable of reasoning about or grasping the relationship between facts. This group passes into the mentally defective.

Class T—Imbecile

According to the second method (hereafter designated, Scale II) each judge makes his own scale. He chooses from among university students whom he has met, five who possess the degree of intelligence denoted by the five steps on the scale. Then he compares each person to be judged, with each member of the scale and assigns him the appropriate numerical value. As heretofore applied, this method has been used in making estimations concerning complex abilities which are subdivided into simpler qualities: ability as an army officer, for example, into physical qualities, intelligence, leadership, personal qualities, general value to the service. Being concerned in the present study only with intelligence, the writer chose this single trait, modeling directions for rating as closely as possible according to the directions for rating army officers. The blank given to instructors is reproduced here, and will make further explanation unnecessary.

RATING SCALE FOR USE IN ESTIMATING INTELLIGENCE

This scale is designed to give basis for the scientific estimation of intelligence, defined as: "the general capacity of an individual consciously to adjust his thinking to new requirements; it is general adaptability to new problems and conditions of life."

Highest	15	Low	6
High	12	Lowest	3
Middle	9		

HOW TO USE THE SCALE

From your acquaintances among university students select five that range from highest to lowest in intelligence.

Place the most intelligent one in class rating 15. Place the least intelligent one in class 3. Select one halfway between these two, of about average intelligence; place him in class 9. Select one who ranks halfway between the middle one and the highest and place him in class 12. Select one who is halfway between the middle and lowest and place him in class 6.

Now take the student whom you are rating and compare him with each of the five in your scale. Give him the number of points following the name of the student he most nearly equals. If he falls between two students in the scale give him a number accordingly (e.g., if between Low and Middle give him 7, 7½ or 8).

POINTS FOR SPECIAL ATTENTION

1. Use as your criterion only General Intelligence, disregarding moral and volitional qualities and scholarship record.
2. In making your scale use students who exhibit clearly the degrees of intelligence assigned to them.
3. The names of the highest and lowest must represent extreme cases.
4. Your scale must be evenly graduated. There should be equal distances between the several steps.

For the purpose of this investigation 100 students were chosen who had been rated the greatest number of times since the adoption of Scale I. Their names were then given to the instructors with the request that they rate the students according to Scale II. It should be stated that the members of the faculty are accustomed to making estimations about students, a systematic plan having been in operation even previous to the adoption of the Pearson Scale; and they responded willingly and whole-heartedly to the request.

Results and Conclusions

TABLE II

	Number of Judgments			Average Scores					Correlations with[5]			
	Min.	Max.	Av.	Min.	Max.	Av.	M. V.	% D.	Grd's	P.E.	Tests	P.E.
Scale I	3	20	10.1	257.0	421.0	331.8	36.0	10	.62	.004	.40	.088
Scale II	1	7	3.0	5.2	13.4	9.9	0.91	10	.66	.012	.32	.094

These scales, so dissimilar to the naked eye, containing different definitions of intelligence, and requiring different techniques in manipulation, nevertheless give results of great similarity, as will be seen from an examination of Table II.

The first thing to strike one's attention is the high average intelligence imputed to this group of university students according to both scales. In terms of Scale I the average for the population-at-large is 300 mentaces; that for these students, 331.8. According to the construction of Scale II the average for *all university students* is 9; that for these students, 9.9. The facts are shown graphically in Graphs I and II, representing the distributions of ratings made by the two scales. The curves both skew toward the upper levels of intelligence. That for scale II is the more nearly normal, probably because the norm taken for the judgments was the university student of medium intelligence, while the norm taken for Scale I was the average intelligence of the population-at-large. The data give good indication that both scales are able to differentiate a university group from a non-selected group.

As to the consistency of the two scales in arranging the members of this selected group with respect to each other, the index of correlation between the two series of ratings is .61, $\pm$ .042, certainly significant, though not remarkably high.

As to the respective values of the two scales in measuring intelligence, we have only indirect light. The only access we have to this elusive trait is that gained through its expressions. Two commonly accepted expressions available for this investigation were academic grades and combined scores made in the psychological tests given to each student in his freshman year.[6] Ratings made with Scale I correlate with grades by .62, $\pm$ .004; ratings made with Scale II correlate with grades by about the same amount. Ratings made with Scale

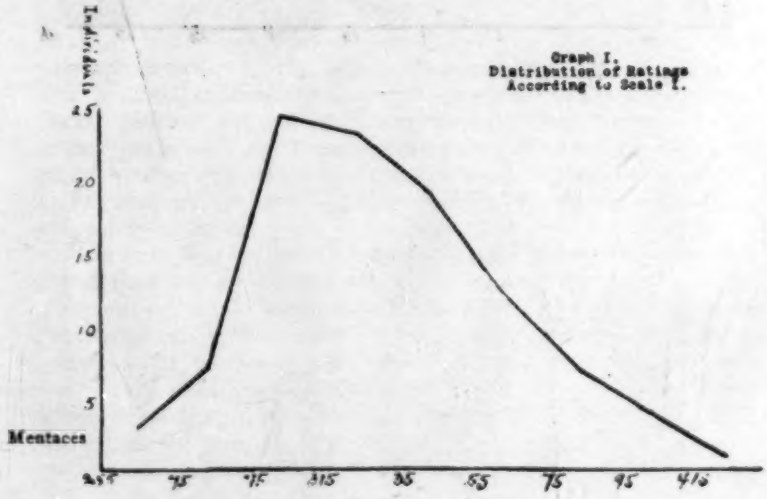
[5] Correlations computed according to formula for adapting the Pearson method to rank-differences.

[6] *loc. cit.*

I correlate with scores made in the psychological tests by .40, ±.088; ratings made with Scale II correlate with test-scores by a closely similar amount. These figures tell us nothing about the *relative* accuracy of the scales in measuring intelligence. They inform us merely that rating by Scale I is about as effective in approximating academic grades as is rating by Scale II; and is about as effective in approximating scores made in the psychological tests as is rating by Scale II.

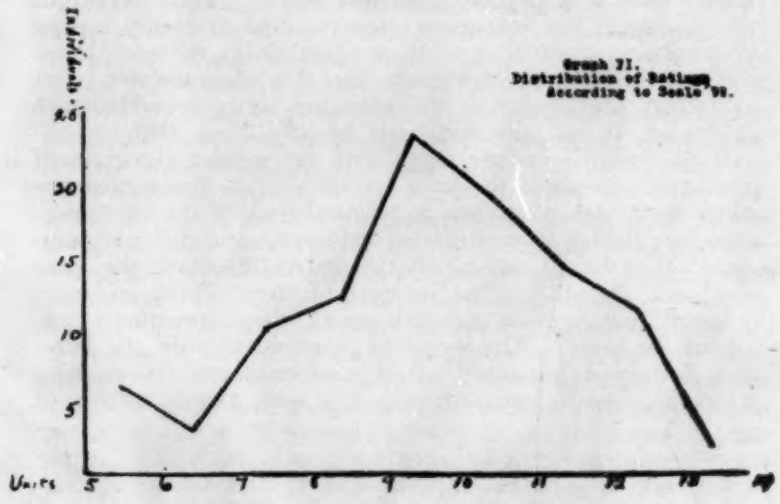
Both series of ratings correlate more closely with academic grades than with psychological test scores. This may argue that in rating, the instructors used the kind of ability leading to scholastic attainment as their chief gauge of intelligence. This conclusion is favored by the fact that when the two series of ratings are combined, the resulting series correlate with grades by .76, ± .028; with tests, by only .39 ± .089.

Both series of ratings agree with the available criteria of intelligence to about the same extent. Does this consistency mean that there is but one thing measured by the two scales even though it is given different definitions and different unit-steps? Or does it mean that the instructors made the same scale out of both? That in spite of their efforts to use a different technique in the two cases, they unwittingly employed the same? The writer is prepared to view the latter as a plausible contingency. It is possible that in making such estimations each judge prefers his own fixed method of



procedure and leans toward it in spite of his efforts to adapt himself to the requirements of one scale or another.

These similarities in the operation of two apparently dissimilar scales suggest that in these days when rating scales are springing up so abundantly, time should be taken for *comparative* investigations. It is quite likely that two methods which at first blush seem widely divergent may turn out upon statistical analysis to be quite similar after all.



Graph 71.
Distribution of Ratings According to Scale 98.

THE AVERAGE MENTAL AGE OF ADULTS

By Edgar A. Doll, Fellow in Psychology, Princeton University

Three years ago Professor Lewis M. Terman came to the conclusion that the average mental age of adults is approximately 16 years. This conclusion is still generally accepted and still determines the calculation of intelligence quotients. Terman based his conclusion on the fact that the median mental age of a fair number of supposedly unselected adults proved to be 16 years by the Stanford revision of the Binet-Simon scale. Terman made no attempt logically to explain the fact nor to elaborate the various implications of his conclusion. Arthur S. Otis, however, by logical analysis and by the application of results obtained by Terman and Kuhlman, showed that the curves of distribution of intelligence at different age groups tended to approach each other more and more closely until the curves of distribution for ages above 16 practically coincided with the 16-year curve. Otis, therefore, concluded for logical reasons, making use of the best data available, that 16 years was the upper limit of intelligence development on the average. Strangely enough, these conclusions aroused comparatively little discussion, but were accepted practically at face value. Many psychologists were surprised that the average intelligence of adults proved to be so low, but those who experimented for themselves, came rather generally to the same conclusion as did Terman. A few isolated studies showed that the intelligence of different social classes and of different occupational groups varied rather definitely, but no attempt was made to obtain any large number of really unselected adults, made up of the representative proportions of the different social levels that are found in society at large.

Psychological examining in the Army, however, has indicated that for unselected adults, as judged by draft recruits, a mental age of 16 years is in all probability much too high for the American average adult, at least as he was observed in the Army. The results of the application of group tests at different Army cantonments indicated wide differences in intelligence according to geographical sections of the country. They indicated very important influences of selections, even in groups of over ten thousand. It was also found that occupa-

tional groups were of strikingly different levels of average intelligence. Furthermore, such studies as were made of foreign-born and negroes, indicated very striking differences according to nationality and race.

Perhaps it is too early to come to any general conclusions regarding the significance of the Army data. Perhaps it is impossible for any one man at any one or even several camps to come to a final conclusion, because of the differences just mentioned in the selected nature of supposedly unselected groups. The fallibility of statistics based even on thousands of cases was amply demonstrated in the Army examining. In spite of highly objective methods, which were almost totally independent of any personal equation of examining and scoring, the different camps reported strikingly different results, and even the same camps reported striking variations. At Camp Dix, for example, the average intelligence scores of successive draft quotas became progressively lower during the period of January to August 1918, although apparently there was no good reason why they should do so. For these and for other reasons, it is advisable to wait until the Surgeon-General of the Army or the National Research Council presents the official results of psychological examining in the Army. The present writer, however, has found so much demand for statements of the Army results, and so much need for revision of former concepts, that he feels it is now very desirable to present some sort of statement, even if it be only preliminary, regarding the average intelligence of adults as seen in the Army. In so doing he will present only very general conclusions. In his own work on return to civil life, the writer has found these conclusions of very practical value, and substantially correct. He confesses, for one, that his own previous opinions of the average intelligence of adults were very much distorted by limited experience and by failure adequately to weight the various experimental groups on which his previous conclusions had been based. The adult population of the State of New Jersey, for example, includes such heavy proportions of foreigners and unskilled laborers that the supposedly unselected groups previously observed by the writer now seem to have been very definitely selected in favor of superior intelligence on the basis of social standing.

The Intelligence of Army Recruits

The following table throws some light on the average intelligence of adults in terms of mental age, and compares native whites, foreign-born, and negroes:

MENTAL AGES OF ARMY RECRUITS BY NATIONALITY AND COLOR

Army Letter-Grade Rating	Equivalent Mental Age	Native White %	Foreign born %	Negro %	Army as Whole* %
Illiterate		17	72	62	30
D—	under 9.5	5	55	39	10
D	9.5—10.9	15	29	28	15
C—	11.0—12.9	21	8	16	20
C	13.0—14.9	27	7	10	25
C+	15.0—16.4	19	1	4	17
B	16.5—17.9	9		2	9
A	18.0—19.5	4		1	4
Total Number		9,500	3,700	3,300	1,500,000
Lower quartile (approx.)		11.5	below 9.5	below 9.5	11.0
Median Mental age (approx.)		13.5	9.5	10.0	13.5
Upper quartile (approx.)		15.5	10.5	12.0	15.5

*Observe that the percentages in this column correspond closely to the weighted average percentages obtained by combining the three preceding columns.

The column headed "equivalent mental age," is obtained from the *Basis for the Assignment of Letter Grades,* Appendix A, of the *Examiner's Guide for Psychological Examining in the Army* (Page 91). The column headed "native whites" is based on approximately 9,500 men of the July draft at Camp Dix, New Jersey. The column headed "foreign-born" is based on approximately 3,700 foreign-born men chiefly southern Italians, Russians and Russian Poles from the July draft at Camp Dix, New Jersey. The data of these two columns are embodied in a report from Lieut. Chas. L. Harlan to the Surgeon-General of the Army, entitled "A Study of Foreign-born Men in the July Draft." The column headed "negro" is based upon approximately 3,300 negroes of the July draft at Camp Dix, New Jersey. The data of this column were obtained from a report to the Surgeon-General entitled "A Study of the Intelligence of Negro Recruits," by Lieut. Edgar A. Doll with the assistance of Lieut. Arthur Kornhauser. The column headed "Army as a whole" is obtained from data presented on page 5 of *Army Mental Tests,* Washington, D. C., November, 1918.

The significance of the table is obvious. The median men-

tal age of the total Army proves to be about 13.5 years. This median presumably includes native whites, foreign-born, and negroes. This median is not substantially different from that obtained for native whites at Camp Dix, but as has been previously mentioned, a median based on any one camp, or on less than many thousands of cases, is not entirely dependable. At Camp Taylor, for example, there were very striking differences between the average intelligence of regiments in the same brigade. We may, however, conclude with a fair degree of safety, that the average of totally unselected adults, regardless of nationality and color, is approximately 13 to 14 years.

The other columns of the table indicate striking differences between native whites, foreign-born, and negroes. The average mental age of native whites is 13.5 years, and of foreign-born and negroes is approximately 10 years.

In considering the average mentality of foreign-born it must be clearly understood that there is no implication that these foreign-born adequately represent the average intelligence of their country. On the contrary it is commonly recognized that the type of immigrant received in this country since 1900 is distinctly inferior to the type of immigrant previously received. In general the foreign-born in this study are probably of inferior intelligence in their own countries. These foreign-born do, however, adequately represent the foreigners now in this country between the ages of 20 and 30.

The column of this table headed "Army letter-grade rating," combines both Alpha ratings and Beta ratings, that is, ratings of both literates and illiterates in group examinations. The first horizontal line of the table shows the percentage of illiteracy and indicates the close relationship between per cent of literacy and average intelligence. The foreign-born and negroes were not penalized because of illiteracy, but rather had the advantage. There was a consensus of opinion among Army examiners that illiterate men of inferior intelligence tended to obtain higher ratings in the illiterate group examination (Beta) than they would have obtained in the literate group examination (Alpha) had they been able to read and write. Compared with the native whites, therefore, the foreign-born and negroes have probably higher ratings than they deserved. On the other hand, it is of course conceivable that illiteracy tends to inhibit the full expression of the innate or potential intelligence of a man who has not had the advantage of schooling and an ordinary social environment. In opposition to this, however, is the conviction of most examiners that

the negroes and foreign-born were adequately rated, and if anything, rated higher than their true intelligence warranted.

We come to the general conclusion, then, that the average intelligence of native whites is approximately 13 years, and that the average intelligence of foreign-born and negroes is approximately 10 years. It should be stated that there are striking differences in the mentality of the different foreign-born groups. The column of this table headed "foreign-born," is based chiefly upon southern Italians, Russians, and Russian-Poles, who constitute approximately 2,500 of the total 3,700. It is not feasible at this time to consider the intelligence of the foreign-born in relation to specific nationality. This was an important contribution in the study made by Lieut. Harlan quoted above. Similarly among the negroes there were striking differences between the northern and southern negroes. The southern negroes were decidedly inferior to the northern negroes. The southern negro had an average mental age of about 9 years as compared with the average mental age of about 11 for northern negroes, as indicated in the study of negroes referred to above.

The Age of Average Mental Arrest

The conclusions drawn from the preceding argument indicate that the average intelligence level of native white adults is approximately 13 years. This implies that growth in general intelligence is complete at the life age of 13 years, but this is not a necessary corollary. Such a conclusion cannot be made until successive examinations over a period of years on large numbers of cases, indicate at what age intelligence actually does cease to develop. It is conceivable that while the average intelligence level of adults is 13 years, this development may not have been fully attained until perhaps considerably after the life age of 13 years. There is indeed some reason to believe in a decreasing rate of intelligence growth, so that complete intelligence development may not be attained until an indefinite age.

Some light is thrown on this problem by applying the Army Alpha test to juveniles. The writer has examined 514 children in the Franklin School, Trenton, N. J., with the Alpha group test. The median Alpha scores at each age are as follows:

Median Age	9.5	10.5	11.5	12.5	13.5	14.5	15.5
Median Score	26	37	61	81	85	80	88
No. Cases	29	38	63	96	139	75	22

These figures indicate that the median intelligence of children above 12 years does not significantly exceed the median of 12-year-old children. There is a slight superiority of 13 over 12, but after this there is a decrease. These children were examined in the eight grammar grades of a typical average public school. An analysis of the age and grade distribution indicates that the children are typically average school children. The children in the 15-year-old group might, however, be considered as slightly below average since they are slightly over age for the eighth grade.

A similar study has been made of 480 delinquent boys at the State Home for Boys at Jamesburg, N. J. The median scores at each age are as follows:

Median Age	9.5	10.5	11.5	12.5	13.5	14.5	15.5	16.5
Median Score	6	8	15	20	31	34	30	29
No. Cases	15	33	51	82	91	102	73	28

It will be observed that at each age these Alpha scores are markedly below the scores of the public school children of the same age. In fact, the average inferiority amounts to about 3 years. The point at issue, however, is that here again average development is complete at about 13 years, since the median scores of ages above 13 do not materially exceed those at 13, and in fact at some ages are below. This substantiates the result found in the public school and indicates that on the average, mental development is complete at about 13 years.

It is not possible to cite any satisfactory statistics for any long period of years on the successive mental ages of normal children of adolescent age. The writer is, however, conducting a study of approximately 250 feeble-minded children who have been re-tested over a period of at least four years each. It is too early to state final results, but as far as the study has progressed it is evident that the feeble-minded do not develop in intelligence after the life age of about 13 years. Contrary to previous beliefs, the age of final arrest is not a function of mental retardation, but is a fairly definite life age. Some feeble-minded of each grade continue to develop in general intelligence, even though only slightly, up to about 13 years of age.

From all of this evidence it appears that not only is 13 years the average intelligence age of adults, but the implication is very strong that the full level of intelligence has been reached at the life age of approximately 13 years. This is not to be understood, however, as meaning that the entire mental growth is attained at that age. Maturity and experience must

certainly continue to effect changes in mentality after the age of 13 years, but the intelligence *level,* on the average, remains fixed. This does not mean that no one develops in intelligence after 13 years of age, or that no one has a mental age above 13 years. On the contrary, 50 per cent of these groups have mental ages above 13 years. It is not possible to say at this time which of the 50 per cent continue to develop, nor to what mental age they develop, nor at what life age limits mental growth ceases. The only permissible conclusion from these data is that ***on the average,*** or for 50 per cent of presumably unselected cases, intelligence growth is practically complete at 13 years.

The writer wishes to emphasize that it would be fatal to both theory and practice to establish anything more than some new hypotheses on the data of this paper. Fifty per cent of these subjects do seem not to have developed intellectually after 13 years, but we cannot say much about the other 50 per cent. Repeated individual examinations on a large scale are now our most pressing need.

Intelligence and Feeble-mindedness

There has been a tendency among clinical psychologists to lay increasing emphasis upon intelligence as a diagnostic symptom in determining feeble-mindedness. This emphasis is entirely proper but has been subject to a certain logical error. It was found by examining mental defectives in institutions that practically no feeble-minded "tested" higher than 13 years in the Binet-Simon scale. It was then concluded both by implication and by actual statement, that all who tested *under* 13 years were feeble-minded. The upper limit of feeble-mindedness was tacitly assumed as the lower limit of normality. Obviously this is the logical error of the converse. If all A is B, the only necessary consequence is, that *some* B is A. For example, if all negroes have mental ages below 15 years it does not follow that all persons with mental ages under 15 years are negroes.

From practical experience, some clinicians have developed the concept of borderlinity. They recognize that there is a certain mental age zone within which the intelligence level alone breaks down as a diagnostic instrument. Within the limits of this zone some persons may be feeble-minded, some doubtful and some normal. In this connection the important contribution of the Army examining lies in the demonstration that persons with mental ages as low as even 7 or 8 years may

be socially normal, and that others with mental ages as high as 13 or 14 years may be socially feeble-minded.

The fundamental conception of feeble-mindedness, of course, is one of social competence. A feeble-minded person is not a person of a particular mental age, but is one who by reason of inferior intelligence is incapable of supporting himself independently of supervision in society, after he has reached the age of social responsibility or physical maturity. The mental age limits which were developed by clinicians in the Army on the basis of practical experience to designate this social responsibility are somewhat as follows:

Diagnosis	Native White	Foreign-born	Colored
	Mental age	Mental age	Mental age
Invariably feebleminded	under 8	under 7	under 7
Usually feeble-minded	8.0— 8.9	7.0— 7.9	7.0— 7.9
Usually inferior normal	9.0—10.9	8.0— 9.9	8.0— 9.9
Usually average normal	11.0—13.9	10.0—11.9	10.0—11.9
Usually superior normal	above 14.0	above 12.0	above 12.0

In this table "invariably feeble-minded" means that persons of the intelligence level specified invariably proved to be feeble-minded when account was also taken of their social success, educability and general mental responsibility. "Usually feeble-minded," means that those persons within the mental age limits specified, usually, or in about 8 cases out of 10, proved to be feeble-minded when other factors are considered. In this group those who were not feeble-minded were ordinarily subnormal or borderline cases. "Usually inferior normal," means that in about 8 cases out of 10 the individual within the mental age limits specified proved to be normal on the basis of social responsibility, literacy, educability and general mental qualities. Those not found to be normal were ordinarily high-grade feeble-minded or borderline cases.

This grouping is not based on statistical or experimental study, but is an empirical classification developed from extensive experience. In the Army it was found impossible to use intelligence alone with any degree of certainty as a diagnostic method above intelligence level of about 8 years. Mental age was an invaluable diagnostic *aid*, however, and was perhaps the most important consideration next to a man's social history. In the Army it was found necessary to weigh social success very heavily in determining normality or feeble-mindedness, and in the mental examinations it was found important to take serious account of the quality of intelligence as well as the

level. Temperamental characteristics were also very important; prudence, forethought, judgment, reason and insight were opposed to mere verbal brightness.

It will be noted that this classification uses lower mental age standards for foreign-born and negroes than for native whites. This implies that the mental ages of foreign-born, negroes, and native whites are of somewhat unequal value in determining social fitness. The mental ages of the foreign-born and negroes seemed to be accurately measured, but due to some unknown psychological quality, the social value of mental age for foreign-born and negroes was somewhat greater than for the same age for native whites.

This empirical conclusion suggests an important problem for experimental psychology. If the prognostic value of mental ages is not constant for the different races, psychology should certainly aim to discover the reason. It seemed desirable in the Army, and still seems desirable, for practical social purposes, to judge individuals according to standards for their race and social class, but this leads to some confusion in the diagnoses obtained. We have come so much to depend upon mental age as an expression of social fitness, that it will be somewhat difficult to employ multiple standards for different social groups.

Average Intelligence and the I. Q.

It is now the practice to use 16 years as the basis of calculation for subjects whose life age is higher than 16. But if the implications here presented should be substantiated by further investigation, then it will become necessary to compute I. Q's. on a 13-year final age. This will make necessary a revision of previous work in which I. Q's. have been stated independently of age. Until the question of the average age of final mental development is settled, it should be urged upon all advocates of the I. Q. always to accompany I. Q's. with the actual ages of the individuals tested. This is important anyway, since I. Q's. are not directly comparable independent of life ages any more than diagnoses by race and class are comparable independently of the race and class standard. An I. Q. of 50 where mental age is 8, cannot be directly compared with an I. Q. of 50 where the mental age is 4.

It will be noted from the intelligence classification given above, that the I. Q. limits of feeble-mindedness are different from those advocated by Terman and his followers. According to this table, the "invariably feeble-minded" are found to have I. Q's. under 50 as opposed to the I. Q. of 70 advocated

by Terman, and the "usually feeble-minded" are found to have I. Qs. between 50 and 55, instead of between 70 and 75, as advocated by Terman. But, if the I. Q's. of the future come to be calculated on the basis of 13.5 years as the average adult intelligence, instead of 16 as now computed, then I. Q's. under 60 will "invariably" denote feeble-mindedness, and I. Q's. under 64 to 70 will "usually" denote feeble-mindedness. Thus we might retain the present I. Q. diagnostic limits advocated by Terman, but they would have to be calculated on a different basis.

Summary

1. The average mental age of adults is found to be approximately 13 years instead of 16 as heretofore believed.

2. The growth of general intelligence is found to be practically complete on the average by 13 years of age and is not on the average thereafter exceeded. This applies strictly to the level of intelligence or degree of brightness as opposed to intelligence plus maturity, experience, and acquisitions.

3. These conclusions are induced by four separate investigations on totally different types of subjects and from different points of view, namely:

(a). The application of Alpha and Beta group intelligence tests to about 1,500,000 soldiers and recruits, where the average mental age is found to be about 13 years.

(b). The application of Alpha group intelligence tests to about 500 typical public school children, where the median scores of ages above 13 do not exceed the median for 13 years.

(c). The application of Alpha group intelligence tests to about 500 juvenile delinquent boys, who are found to be of inferior intelligence but whose level of intelligence is fully attained by 13 years, on the average.

(d). The repeated application of Binet intelligence tests to about 250 feeble-minded persons of wide ranges of life age and mental age, who individually may develop in intelligence up to a life age limit of 13 years and not thereafter.

4. The conclusions are induced not merely by the empirical data of each separate investigation, but particularly by the consistency of agreement obtained from such strikingly different methods of approach.

5. Striking differences are observed between the average levels of general intelligence of native whites, foreign-born, and negroes. Striking differences also are found for different social classes of native whites, different nationalities of foreign-born, and different geographical groups of negroes.

6. A logical fallacy is exposed in the application of mental age limits to the diagnosis of normality and feeble-mindedness. The conception of borderlinity in relation to mental age is much extended.

7. The calculation of intelligence quotients is found to be seriously disturbed by the uncertainty of the exact life age at which intelligence development is to be considered complete on the average.

8. The standardization of mental age scores for ages above 13 years by any method other than the percentile scores is not attempted, but is found to be impeded by the difficulty of obtaining fully representative unselected groups of subjects outside the grammar schools.

Addendum

It is now possible to strengthen the above argument still further. Nearly 550 delinquent youths and young men between the ages of 16 and 29 inclusive, inmates of the New Jersey Reformatory, have recently been examined by Mr. J. W. Ellis, Assistant Psychologist to the N. J. Dept. of Institutions and Agencies, by means of the Army group test Alpha. The median Alpha scores at each age are as follows:

Median age	16	17	18	19	20	21	22	23	24	25	26	27	28	29	Total
Median score	59	62	56	58	57	43	28	69	38	39	60	98	22	38	56
No. cases	68	115	109	74	47	40	17	15	8	11	11	6	8	5	541

It is obvious from this table that there is no consistent increase in score with increase of age. During the range of years where there is a sufficiently large number of cases for the median to be dependable (16 to 21) there is no significant variation in score except for an unaccountable drop at year 21. The variations beyond year 21 may be accounted for on the basis of selection and the small number of cases. The fact that there is no consistent trend in these variations is, however, more significant than the variations themselves. The median score for the group represents a median mental age of approximately 14 years. The group is a selected group on the basis of delinquency yet the average intelligence score is nearly the same (slightly higher) than the average score of army recruits. Here again, however, we are not so much interested in the absolute score as the fact that there is no steady increase in score within increase in age. The intelligence levels above 16 are not superior to that of 16. It is probable, in the light of the public school and the Jamesburg Home

median scores that the 16-year median itself is perhaps not greater than it would have been had we had age groups extending down to 13 or 14 years, but this is conjectural.

This reformatory group supplies a fifth group and a new type of subject. The group covers the age-period between the public school group and the Army adult group. It fills the gap and gives a continous range of ages from 6 to 31 years. (The scores for years 6, 7 and 8 years have been omitted in this paper.) The continuity is, of course, broken by the variations in types of subjects, but an adequate basis of preliminary argument is afforded in the actual results. The new data confirm the preceding argument that there is no increase in average general intelligence scores after a chronological age of presumably 13 or 14 years. Again we must insist that the scores of individuals range far above the 13-year level, with the presumption, however, that such scores represent superior deviations of the 13-year group and not age-increases as such. These superior scores also demonstrate that the 13-year level is not caused by limitations in the scale of tests itself, since individuals do score far beyond the median, thus proving the practical possibility of obtaining high scores if the capacity is present in the individual.

A PSYCHOLOGICAL INVESTIGATION OF THE LIKELIHOOD OF CONFUSION BETWEEN THE WORDS "COCA-COLA" AND "CHERO-COLA"[1]

By Richard H. Paynter, Jr.

This is a report of a series of psychological experiments conducted to investigate the desceptive similarity or likelihood of confusion in visual appearance and in sound between the word trade-marks "Coca-Cola" and "Chero-Cola." The report also compares the likelihood of confusion between these words with that found to exist between word trade-marks that have been held in judicial proceedings to conflict and with others that have been held not to conflict. Two distinct methods were employed in the investigation: the recognition or identification method, and the method of relative position. Four separate experiments were conducted, three with recognition and one with relative position. The recognition experiments were on confusion in visual appearance, and the relative position experiment was on confusion in sound.

A. Visual Recognitive Confusion

1. *General Procedure and Technique.*—The visual recognition or identification method was used because it involves the same mental process as that employed in daily life by the ordinary purchaser in buying a bottle of "Coca-Cola." In ordinary transactions of business where the article is sold across the counter, the purchaser examines the article to see if it is what he wants, or whether it is the same as that which

[1]This study is a practical application of simple, straightforward psychological methods to an actual trade-mark case. The study was used in a case by The Coca-Cola Company in an opposition to the word "Chero-Cola."

A more extensive investigation will appear in a forthcoming number of the Archives of Psychology, entitled "A Psychological Study of Trade-mark Infringement."

The writer expresses his deep gratitude to Professor Edward K. Strong, Jr., of the Carnegie Institute for Technology, for the kind advice and help he has given in devising the experiments.

he has known or bought before. In deciding this matter, the purchaser is guided by the recognition or identification of the article as a whole, by certain features of it, or by a combination of both these methods. It often happens, however, that a prospective purchaser has heard or seen only the word trade-mark, but not the article itself nor its container. In such cases he has to rely on his memory of just the word trade-mark in identifying the article. Likewise, the observer or individual in the experiment has to rely on his memory of the word trade-mark in order to tell whether or not what is before him now is the original trade-mark that he has just seen. The task assigned to him in the experiment was simply to indicate whether the words before him were just shown to him.

The individual who submitted himself to do the recognition experiment was shown in the presentation 20 slips of paper on each of which was typewritten a word trade-mark, or a word trade-mark and the name of its commodity. These slips were presented one at a time at the uniform rate of one a second, by the writer who conducted all the experiments. Fifteen seconds after the exposure of the last slip in the presentation, the test set of slips was given to the observer. The test consisted of 40 slips, and from it the observer was asked to pick out those that he recognized as having just been seen and those that had not been seen. There were two variations of the formation of the test, according to the problem of the experiment. When the problem was to determine the confusion between the words "Coca-Cola" and "Chero-Cola," "Coca-Cola" appeared in the presentation, whereas the imitation "Chero-Cola" appeared in the test together with 19 trade-marks which were duplicates of 19 shown in the presentation, and 20 trade-marks which were new, i. e., they had not been shown in the presentation. When the problem was to compare the confusion of the imitation "Chero-Cola" with that of 9 other litigated imitations, "Coca-Cola" with their 9 originals appeared in the presentation, whereas "Chero-Cola" appeared in the test together with these 9 imitations, 10 duplicates of the remaining 10 trade-marks shown in the presentation, and 20 new marks. In all these experiments "Coca-Cola" was shown in the presentation, but in the test where it did not appear the observer had to decide whether he had seen "Chero-Cola" which was not shown in the presentation. If he decided that he had seen the imitation he was confusing it with the imitation; if he decided that he had not

seen the imitation he was not confusing it with the original.[2] Where the comparison was made with other litigated imitations the name of the article or commodity appeared with each trade-mark. There are two experiments where no comparison was made, one showing the trade-marks without, and the other with, the names of their articles. It should be stated here, however, that although the words in the recognition experiments entered the mind of the observer through vision, not only did similarity in visual appearance, but also in sound, linguistic formation, and meaning or significance contribute to confusion.

The following directions were given to each observer before he did the experiment:

"You are going to be shown one at a time a number of ordinary word trade-marks like 'Uneeda,' 'Garford,' and 'Celluloid.'" (In the experiments where the trade-marks were used with the names of the articles, the observer was told of this and the examples were given as "Uneeda Biscuit," "Garford Automobile," etc.) "You are to read all the words on each slip of paper. Read them naturally as though you were reading an advertisement in a magazine or in a street car. Fifteen seconds after you have been shown the last mark, you will be given a second set (the test) and asked to pick out those marks you have just seen in the presentation and those which you have not seen. You will be further asked to sort the marks into six piles, according to the degree of your confidence or certainty of your recognition of your marks. There are three degrees of certainty for the marks that are recognized as seen, and three similar degrees for those that are recognized as not seen. The three degrees are 'absolutely certain,' 'reasonably certain,' and a 'faint idea.'" (After the observer has indicated his recognitions of all the marks in the test, he was asked how familiar he had been with the word "Coca-Cola"). "Would you say you had been 'perfectly familiar,' 'moderately familiar,' 'just familiar,' or

[2] These two statements are not strictly accurate as they do not take into consideration the possibility of the presence of a small per cent. of observers who would not have been able to recognize "Coca-Cola," and in whom the perception of "Chero-Cola" would not recall "Coca-Cola." These observers would be guessing in recognizing whether or not they had seen "Chero-Cola," and would guess correctly in about 50 or 60 per cent. of the cases, thereby influencing equally the correct and incorrect recognitions of "Chero-Cola." As the 10 original trade-marks are approximately equally well recognized or familiar, accounting for this factor would introduce no significant change in the data or conclusions.

'unfamiliar' with the word 'Coca-Cola' before you did this experiment?"

Thus, three important facts were obtained from the experiment on each observer. The first fact was, whether or not he confused "Chero-Cola' with "Coca-Cola." The second fact was, how confident he was that he was correct, or whether he was in doubt about his recognition. And the third fact was, how familiar he had been in daily life with the word "Coca-Cola." The question of visual confusion will be determined then by the data from these three sources of information.

Each observer did only one experiment; and he did not do the same experiment more than once. The observer was not informed of the purpose of the experiment, nor did he know beforehand that an original trade-mark would be replaced by an imitation in the test. Furthermore, nothing was done to cause the observer to have any suspicion that a substitution had been made. The task set the observer in the experiment was not arduous, nor unpleasant. The experiment was done with the observer's own free will. No prize or reward was offered for the highest individual record. After each observer did the experiment the marks in both the presentation and test were thoroughly shuffled. As their positions in both series were thus determined by chance, no trade-mark was given undue prominence by its position.

Ninety ordinary individuals were employed as observers in the recognition experiments. They represented more than 60 different professions and occupations. Their ages were generally between 20 and 45. These individuals represent a random picking from ordinary individuals that resided in New York City, in the month of July, 1915. No effort was made to obtain exceptionally bright nor exceptionally dull individuals.

All the trade-marks or trade-names studied in the experiments are used or have been used in commerce. When the names of the articles are used in the experiments in connection with the trade-marks, they are the names of the same articles or commodities on which the trade-marks have actually been used. The names of the goods represent many classes of goods of various descriptive properties. In the presentation no trade-mark nor name of a commodity was the same as any other. The names of the commodities of the original trade-marks and their respective imitations were identical. The 20 new trade-marks and the names of their commodities were all different from the others, either in the test or in the presentation.

The trade-marks and names of the commodities were all typewritten in the same kind of type. The experiments did not reproduce the style, coloring, or size of the trade marks or names of the commodities as used on the goods or in advertisements. The color of the ink used was black; and great care was taken in typewriting the words to keep the blackness of the ink the same for all the letters. All the words were typewritten on the same kind of white paper. Each slip of paper was two and three-quarters by four and one quarter inches in size. When the name of the commodity was used it appeared on the next line directly beneath the trade-mark. All the words appeared in the second horizontal quarter from the top of the slip. The four following reproductions are samples of some of the slips, which are identical with those actually used, with the exception of being reduced one-fourth.

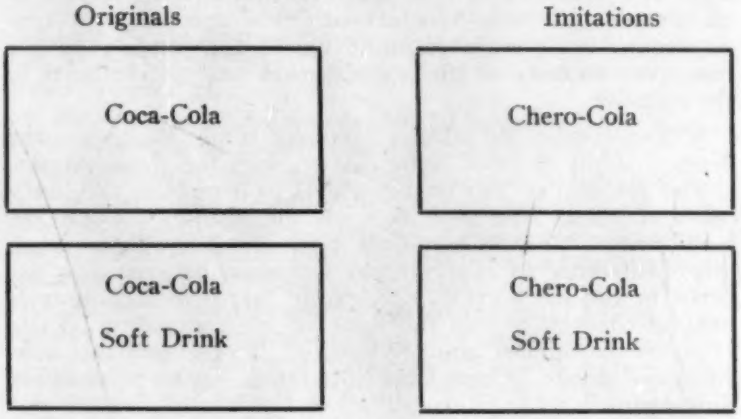


FIG. 1

2. *Results.* a. Experiment in which Appeared One Imitation, the Names of the Articles not being Shown with the Trade-Marks.

The word "Coca-Cola" and 19 other trade-marks appeared in the presentation series; the test series included the word "Chero-Cola," 19 duplicate trade-marks and 20 new trade-marks.

TABLE I

THE NUMBER AND PER CENT. OF 40 OBSERVERS THAT CONFUSED THE WORD "CHERO-COLA" WITH "COCA-COLA"

Number of Observers	Trade-Marks		Confused	
	Original	Imitation	Number	Per Cent
40	Coca-Cola	Chero-Cola	11	28

According to Table I, approximately one out of every three observers or 28 per cent. actually takes the word "Chero-Cola" to be "Coca-Cola." Of the 11 deceived by "Chero-Cola" 9 were "absolutely certain," 1 "reasonably certain," and 1 had a "faint idea" that they had seen it. Thus most of the observers mistaken were quite confident that they were correct. Moreover, 9 of those confused by "Chero-Cola" were "perfectly familiar," and 2 "moderately familiar" with the original trade-mark "Coca-Cola." It is seen from this that observers even very familiar with the word "Coca-Cola" were deceived by the imitation "Chero-Cola."

Throughout this investigation the figure representing the per cent. of individuals in a group confused by an imitation is not absolute. That is, it does not mean that just this per cent. of observers under the multitudinous varying conditions of everyday life will be confused. The percentage of confusion, is, however, a scientific measure depending on the observer's memory of the original mark and its similarity to the imitative.

b. Experiment in which Appeared One Imitation, the Names of the Articles being Shown with the Trade-Marks.

The problem of this experiment is to determine the likelihood of visual confusion between the phrases "Coca-Cola Soft Drink" and "Chero-Cola Soft Drink." There is no other difference in experimental technique between this experiment and the previous one, except that here each mark is used in connection with the name of its commodity. Table II. gives the number and per cent. of 25 observers that confused the words "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink."

TABLE II

THE NUMBER AND PER CENT. OF 25 OBSERVERS THAT CONFUSED THE WORD "CHERO-COLA" WITH "COCA-COLA," WHEN APPLIED TO THE WORDS "SOFT DRINK"

Number of Observers	Trade-Mark		Name of Article	Confused	
	Original	Imitation		Number	Per Cent
25	Coca-Cola	Chero-Cola	Soft Drink	17	68

Seventeen or 68 per cent. of the group confused "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink." When used in connection with the words "Soft Drink" 2 out of every 3 observers take "Chero-Cola" for "Coca-Cola." Of the 17 deceived by "Chero-Cola" 12 were "absolutely certain," 2 "reasonably certain," and 3 had a "faint idea" that they had seen it. Moreover, 14 of those confused were "perfectly familiar," 2 "moderately familiar," and 1 "just familiar" with

the original mark "Coca-Cola." As in Table I. the observers in Table II. were nearly all quite confident of their false recognitions of the imitation, and also very familiar with the "Coca-Cola."

The figures in Table II. show about two and one-half times as much confusion as do those in Table I. It is the addition of the name of the commodity to both the original and imitative marks that furnishes an explanation for this increase. As the observers in the present experiment had to read in the presentation the words "Coca-Cola Soft Drink" in the same time that those in the previous experiment had to read just the word "Coca-Cola," the word "Coca-Cola" in the former not being so well perceived, was not so firmly established as in the minds of those of the latter group. Having to contend with a weaker mental impression of "Coca-Cola" confusion with "Chero-Cola" was more likely. Furthermore, as the two perceptions "Coca-Cola Soft Drink" and "Chero-Cola Soft Drink" are relatively more similar than the two perceptions "Coca-Cola" and "Chero-Cola," confusion between the former is hence greater.

c. Experiment in which Appeared 10 Imitations, the Names of the Articles being shown with the Trade-Marks.

The problem in the present experiment is to compare the likelihood of visual confusion between the words "Coca-Cola Soft Drink" and "Chero-Cola Soft Drink," with the likelihood of visual confusion found to exist between other pairs of similar word trade-marks that had been the subject of actual adjudication of infringement or non-infringement. Nine legal decisions [3] on the similarity of word trade-marks were selected

[3] Holeproof, Knotair, hosiery; Holeproof Hosiery Co. v. Wallach Bros., 190 Fed. Rep. 606 (U. S. Cir. Ct.).

Grape-Nuts, Grain-Hearts, cereal food; Postum Cereal Food Co., Ltd., v. American Health Food Co., 119 Fed. Rep. 848 (U. S. Cir. Ct. of App.).

Cascarets, Castorets, remedy; Sterling Remedy Co. v. Spermine Medical Co., 112 Fed. Rep. 1000 (U. S. Cir. Ct. of App.).

Gold Dust, Gold Drop, washing powder; N. K. Fairbank Co. v. Luckel, King & Cake Soap Co., 102 Fed. Rep. 327 (U. S. Cir. Ct. of App.).

Listerine, Listogen, antiseptic; Lambert Pharmacal Co. v. Bolton Chemical Corporation, 5 Trade-Mark Rep. 38 (U. S. Dist. Ct.).

Mother's, Grand-Ma's, waxing pad; E. A. Bromund Co. v. Columbia Wax Products Co., 200 O. G. 1115 (Ct. of App., D. C.).

Sozodont, Kalodont, tooth paste; K. K. Landespriv et al. v. Hall & Ruckel, 1911 C. D. 329, 165 O. G. 732 (Ct. of App., D. C.).

Green River, Green Ribbon, whisky; Lang v. Green River Distilling Co., 1909 C. D. 476, 148 O. G. 280, 33 App. D. C. 506 (Ct. of App. D. C.).

Club, Chancelor Club, cocktails; In re S. C. Herbst Importing Co., 1908 C. D. 383, 134 O. G. 1565, 30 App. D. C. 297 (Ct. of App. D. C.).

for comparison. Five were adjudications of infringement, in which the use of the imitative marks was enjoined, and four were adjudications of non-infringements in which injunctions against the use of the alleged illegal imitations were refused. In selecting these decisions two principles were followed; first, that the original mark in the decision should be widely known; and second, that no other circumstance or reason than that of similarity between the word trade-marks should have determined the point of infringement or non-infringement in the decision.

The first principle was set up because the word "Coca-Cola" is widely known, and is pretty familiar to a great many individuals in New York City, where all the experiments were conducted. And furthermore, because an imitation would probably cause less confusion in the case of a more familiar original mark than in the case of a less familiar original, a comparison of the confusion caused by their respective imitations would unfairly show relatively greater confusion for the imitation of the less familiar original. Thus, the confusion brought about by the imitation of "Coca-Cola" will be properly compared with the confusion brought about by the infringing and non-infringing imitations, if their originals are approximately as well-known as "Coca-Cola." It should be stated here, however, that not all the decisions studied have originals as well known-as "Coca-Cola." These marks will receive separate treatment.

The second principle was set up for two reasons. The first was that, the experiment measures visual confusion of just word trade-marks. The second was that, if various factors such as the question of the validity of the trade-marks, unclean hands, similarity of the type, color, or other features of the label or package entered in the decisions and operated to influence the point of infringement or non-infringement, we could not properly compare the experimental results of these complicated decisions with each other, nor with the simple decisions of confusion of just the word trade-marks, nor with "Chero-Cola." It is obvious that decisions determined by confusion of word trade-marks plus certain other reasons are not in the same legal nor psychological categories as those determined simply by the confusion of just the word trade-marks. The decisions in the former categories could not be properly given as authorities governing those in the latter, or vice versa.

Notwithstanding the fact that some selection has been made in obtaining these decisions, no selection was made to obtain

decisions whose trade-marks seemed to have only little likelihood of confusion or less than that of the imitation of "Coca-Cola." That is, no judgment was exercised to set forth as very great the confusion caused by "Chero-Cola" by contrast with imitations that might be expected to show only a slight degree of confusion. In this respect the decisions represent a chance or random sampling of infringements and non-infringements, affording thereby a fair comparison of "Chero-Cola" with infringements and non-infringements of well-known original trade-marks in general. In this respect the experimental method gives a scientific comparison, not one biased in advance to argue a decision of infringement against the word "Chero-Cola."

This experiment is the same as the previous one, except for the presence of 10 imitations in the test series instead of one. The words "Chero-Cola Soft Drink" were in the test series together with 9 other imitations, 10 duplicates, and 20 new trade-marks.

TABLE III

THE NUMBER AND PER CENT OF OBSERVERS CONFUSED BY THE WORD "CHERO-COLA," EACH OF THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS, WHEN APPLIED TO THE NAMES OF THEIR ARTICLES

Decision	Trade-Marks Original	Trade-Marks Imitative	Name of Article	Confused No.	Confused P.C.
N	Sozodont	Kalodont	Tooth Paste	24	48
I	Green River	Green Ribbon	Whisky	23	46
N	Mother's	Grand-Ma's	Waxing Pad	20	40
	Coca-Cola	*Chero-Cola*	*Soft Drink*	*10*	*40**
I	Club	Chancellor Club	Cocktails	17	34
I	Listerine	Listogen	Antiseptic	16	32
I	Cascarets	Castorets	Remedy	14	28
I	Gold Dust	Gold Drop	Washing Powder	13	26
N	Holeproof	Knotair	Hosiery	12	24
N	Grape-Nuts	Grain-Hearts	Cereal Food	9	18

*This percentage is reckoned on the basis of 25 observers in the group; all the other percentages are reckoned on the basis of 50 in the group.

Table III. gives the number and per cent. of observers confused by each of the imitations. The first column in the table shows whether the imitation in the decision was held to be an infringement (I) or a non-infringement (N). The second column gives the original trade-mark in the decision; the third the imitative trade-mark; and the fourth the name of the commodity to which both the original and imitative trade-

marks were applied. The fifth column gives the number of observers confused; the sixth and the last per cent. confused. The marks are arranged in an order of confusion from most to least.

"Sozodont—Kalodont," a non-infringement case, is the most confusing pair, 24 observers or 48 per cent. of the entire group were confused. "Green River—Green Ribbon," an infringement, stands second highest with 46 per cent. confused. "Coca-Cola—Chero-Cola" ties with a non-infringement, "Mother's—Grand-Ma's," for the third most confusing position with 40 per cent. confused. The differences between these four pairs of marks are not large. Of the 10 observers confused by "Chero-Cola" 9 were "absolutely certain," and 1 had a "faint idea" that they had seen it. Nine of those confused by "Chero-Cola" were "perfectly familiar," and 1 "moderately familiar" with the word "Coca-Cola." As in the two previous experiments most of the observers confused were very familiar with the word "Coca-Cola," and they were scarcely even doubtful in their errors. Two imitations are more confusing than "Chero-Cola," one of these is a non-infringing imitation and the other an infringing. Four infringements and two non-infringements have lower scores than "Coca-Cola—Chero-Cola." It is 6 per cent. lower than the most confusing infringement "Green River—Green Ribbon;" and 14 per cent. higher than the least confusing infringement "Gold Dust—Gold Drop." It is 8 per cent. lower than the most confusing non-infringement "Sozodont—Kalodont;" and 22 per cent. higher than the lowest non-infringement "Grape-Nuts—Grain-Hearts."

The original marks in the table that are about as well-known in New York City as "Coco-Cola" are perhaps "Gold Dust," "Grape-Nuts," and "Cascarets." The remaining original marks, with the exception of "Sozodont," contain ordinary words which are met with frequently enough in daily life to be perhaps as familiar as the three above marks. It seems that the combination of the two ordinary words in "Holeproof" puts it also under the latter class of marks. Let us now compare the scores of the imitations of these two classes with the score of "Chero-Cola." It is more confusing than the imitation of any of the most widely known original marks. Two of these imitations are infringements and one a non-infringement. In order of amount of confusion these marks are:

Coca-Cola	*Chero-Cola*	*40*	per	cent.
Cascarets	Castorets	28	"	"
Gold Dust	Gold Drop	26	"	"
Grape-Nuts	Grain-Hearts	18	"	"

" Coca-Cola—Chero-Cola " has confused 12 per cent. more observers than " Cascarets—Castorets;" 14 per cent. more than " Gold Dust—Gold Drop;" and 22 per cent. more than " Grape-Nuts—Grain Hearts." Of the less well-known original marks one non-infringement " Sozodont—Kalodont," and one infringement " Green River—Green Ribbon " are more confusing. " Mother's—Grand-Ma's," a non-infringement ties with " Coca-Cola—Chero-Cola," and all the rest are less confusing than it.

It will be observed that in Table III " Chero-Cola " shows less confusion than in Table II. In both experiments the marks were used in connection with the name of the commodity. In Table III the per cent. confused is 40, and in Table II, 68 per cent., the difference being 28 per cent. The reason for the decrease in Table III was that " Chero-Cola " appeared in the test together with 9 other imitations. The fact that the observer was usually confronted with several imitations before the imitation of " Coco-Cola " appeared, was in itself a kind of fore-knowledge and warning to beware of imitations. In the experiment of Table II. the presence of one imitation among 39 different trade-marks in the test, 19 being duplicates of what they had seen and 20 new, would find the observer not expecting to be fooled by an imitation. In the experiment of Table III the presence of 10 imitations in the test together with 10 duplicate and 20 new marks would, on the contrary, act as a warning to be more cautious, wary, and careful in recognizing all marks. The natural result followed this condition; that is, there was a cutting down of the confusion caused by " Chero-Cola."

TABLE IV

THE AVERAGE PER CENT. OF OBSERVERS CONFUSED BY THE WORD "CHERO-COLA," THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS, WHEN APPLIED TO THE NAMES OF THEIR ARTICLES

Number of Decisions	Decision	Confused Av. P.C.
	Coca-Cola—Chero-Cola	*40.*
5	Infringement	33.2
4	Non-infringement	32.5

If we compute the average percentages of the five infringing marks, and of the four non-infringing marks, as in Table IV, we find that the score of "Coca-Cola—Chero-Cola" is higher than either average. It is about 7 per cent. higher. The average for the infringements is 33.2 per cent., and for the non-infringements 32.5 per cent. The average of the infringements is only 0.7 per cent. higher than the average of the non-infringements.

The most important characteristic of this difference is its small amount. On the other hand, "Green River—Green Ribbon" and "Gold Dust—Gold Drop" is 20 per cent; and the largest difference between two non-infringements, "Sozodont—Kalodont" and "Grape-Nuts—Grain-Hearts" is 30 per cent. Returning for a moment to Table III, two non-infringements, three if we include "Coca-Cola—Chero-Cola," are more confusing than four infringements. The most confusing imitation is a non-infringement; and three of the four most confusing imitations are non-infringements, including "Coca-Cola—Chero-Cola." The least two confusing imitations are non-infringements. The differences within both classes of decisions so enormously outweigh the difference between the classes, that for practical purposes the difference in confusion between the infringing and non-infringing imitations may be disregarded.

Therefore, these decisions are not entirely reliable or consistent. The experiment, presenting conditions no more difficult or essentially different from those in daily life, finds that there is a likelihood of confusion with all non-infringing imitations which the courts held to show no confusion.

d. Duplicate and New Trade-Marks. By calculating how well the duplicate and new trade-marks were recognized, the difficulty of the task set the observer may be shown. If the observer recognizes correctly all the duplicates and new trade-marks, he would then obtain for them 100 per cent. of correct recognitions, or a perfect score. A score of 50 per cent. of correct recognitions may indicate that the experiment is exceedingly difficult or that the observer was simply guessing, in either case meaning zero or no memory of the duplicate shown in the presentation. By pure chance alone it would be possible to obtain this score. An average of 75 per cent. of correct recognitions would then stand midway between zero ability to recognize and just a perfect score; an average of 85 per cent. would stand seven-tenths of the distance to the latter. The other intervening scores can likewise be judged as to their relative difficulty. So, irrespective of the observer's recog-

nitions of the imitations, the difficulty of the experimental conditions may be shown by calculating the per cent. of correct recognitions of the duplicate and new trade-marks.

As is shown in Table V, the duplicates have a score of 80 per cent., and the new trade-marks a score of 94 per cent.

TABLE V

THE AVERAGE PER CENT. OF OBSERVERS THAT CORRECTLY RECOGNIZED THE DUPLICATE AND NEW MARKS, WITH ONE IMITATION WITHOUT THE NAME OF THE ARTICLE

Number of Marks	Kind of Mark	Average Per Cent
19	Duplicate	80
20	New	94
	Average	87

It is thus easier to tell the marks that have not been seen before than those that have. The final average 87 per cent. represents the difficulty of the task set the observers. That is, excluding the results of the imitation, the average per cent. of correct recognitions of all the other marks is 87 per cent. The task was perhaps not more difficult than that which the prevailing conditions in daily life would offer, nor so easy that it could be reacted to without error.

TABLE VI

THE AVERAGE PER CENT. OF OBSERVERS THAT CORRECTLY RECOGNIZED THE DUPLICATE AND NEW MARKS, WITH ONE IMITATION WITH THE NAME OF THE ARTICLE

Number of Marks	Kind of Mark	Average Per Cent.
19	Duplicate	66
20	New	91
	Average	79

Table VI presents the results of the duplicate and new trade-marks in the experiment in which one imitation appeared, and in which the trade-marks were used with the names of their commodities. The duplicates have a score of 66 per cent., and the new trade-marks a score of 91 per cent. In Table VI, as also in the preceding table, the average per cent. of the new trade-marks is higher than that of the duplicates. There is, however, in Table VI a lower score for the duplicates than in Table V; those for the new trade-marks are about the same.

This decrease in per cent. of correct recognitions for the duplicates with the name of the article is explained by the fact that in the same length of exposure more material in the presentation had to be taken in than when the trade-mark alone was shown. A weaker mental impression for the duplicates in Table VI thus caused a drop in the per cent of correct recognitions. But, adding the name of the article to the new trade-marks made just a slightly lower change in ability to discriminate them. We have already seen above in Table II that the relatively shorter time to receive the impression of the original with the name of the commodity aided in causing more confusion than did the original trade-mark without it, as in Table I. The final average in Table VI is 79 per cent. Although it is 8 per cent. lower than the final average in Table V, the task is still of fair difficulty.

TABLE VII

THE AVERAGE PER CENT. OF OBSERVERS THAT CORRECTLY RECOGNIZED THE DUPLICATE AND NEW MARKS, WITH 10 IMITATIONS WITH THE NAME OF THE ARTICLE

Number of Marks	Kind of Mark	Average Per Cent.
10	Duplicate	68
20	New	94
	Average[4]	85

Table VII gives the average per cent. of correct recognitions of the duplicate and new trade-marks in the experiment in which 10 imitations appeared, and in which the trade-marks were used with the names of their commodity. The duplicates have a score of 68 per cent., and the new trade-marks a score of 94 per cent. As in Tables V and VI the average of the new trade-marks is higher than the average of the duplicates. The scores in Tables VI and VII are for both kinds of marks about the same. The final average 85 per cent. in Table VII represents the difficulty of the task.

B. SOUND CONFUSION MEASURED BY RELATIVE POSITION

1. *General Procedure and Technique.* The method of relative position was used to compare the likelihood of confusion in sound between the words "Coca-Cola" and "Chero-Cola" with that between the litigated trade-marks of the previous

[4] This average was obtained by dividing the sum of the 30 separate scores by 30: it is slightly higher than the average obtained by dividing the sum of the average of the duplicate and new scores by 2.

experiment. This method does not state "how many" are confused in sound by the various imitations, but it does give a measure of their relative differences in sound confusion. It required the observer to arrange the 10 pairs of litigated trade-marks in a list according to the amount or magnitude of confusion that the imitation is likely to cause. Any pair is then measured by its position in the list. In the comparison of "Chero-Cola" with the nine imitations, and the psychological criticism of their legal decisions it is recognized that the decisions were not rendered entirely on sound confusion.

Measurement by relative position was adopted: first, because it is a different method from that used in the previous experiment; second, because there is a certain analogy between the mental processes of the observer judging relative differences and the mental processes in the minds of the court judging the question of infringement or non-infringement; and third, because confusion between the words "Coca-Cola" and "Chero-Cola" is especially likely to occur under certain actual business conditions in respect to their sounds.

As regards the first point, it would obviously be of much importance if a method different from that employed in visual recognitive confusion should show that confusion in sound between the words "Coca-Cola" and "Chero-Cola" was above that of the average of the infringing and non-infringing marks. Confirmation of greater likelihood of confusion than infringements and non-infringements by two different methods is certainly significant.

As regards the second point, the mental processes of the observer comparing relative differences in sound confusion are essentially very similar to those in the minds of judges comparing the sound confusion of a pair of trade-marks in a case in court with that of trade-marks adjudicated infringement and non-infringement. Although there are resemblances between the mental conditions that give the experimental data and those that give the legal decisions, there are differences between this experimental method and the legal procedure that have greater significance. With measurement by relative position the accuracy or mathematical validity is higher. In employing a far greater number of observers the arrangement is much less likely to be the result of chance bias of a few individuals. In the average of a large number of judgments a chance bias in any one direction is cancelled, and the result represents the general tendency of all the observers.

Furthermore, the observers are entirely disinterested in the outcome of the experiment. They do not know which trade-

marks have been adjudicated infringement or non-infringement, nor that any pair of trade-marks is being especially investigated. Not knowing that the experiment is being conducted for a pending case the judgments of the observers are consequently without prejudice to either plaintiff or defendant. On the other hand, with legal procedure not more than a few judges pass on the decision; and their judgments tend to some extent to be influenced by the abilities of contending counsels to magnify the differences and increase the similarities of the trade-marks. Measurement by relative position gives an exact measure, whereas a judicial decision throws a case into one or two categories, ill-defined and without quantitative significance. It is thus clear that this experimental method is far superior to the present legal procedure.

As regards the last point, the probability of confusion in sound or pronunciation between the words " Coca-Cola " and " Chero-Cola " is considered because it is in this respect that a purchaser asking for a glass of " Coca-Cola " from the dispenser of the drink at the fountain is likely to be fooled. The purchaser under these conditions does not usually see the label on the bottle or its crown, and and he can not identify the drink by its trade-mark. Assuming that the dispenser is acting honestly in listening to his customer's request, and in fulfilling what he believes to be that request, confusion between the words " Coca-Cola " and " Chero-Cola " is most likely in sound or pronunciation alone.

Each observer was given an envelope and a sheet of directions. Enclosed in the envelope were 10 slips of white paper, on each of which appeared two trade-marks. The slips of paper were the same size as those of the previous experiments; and the trade-marks were all typewritten in black ink. The name of the article or commodity did not appear with the trade-mark. The following reproduction is a sample of one of the slips and is identical with that actually used.

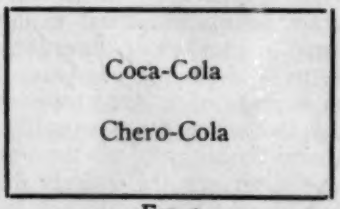


FIG. 2

The directions were as follows:

"Please arrange in serial order as well as you can the enclosed slips of paper according to the following instructions.

"Each slip contains two trade-marks of a common article. The upper is the original trade-mark, and the lower is the imitation of it.

"The different pairs of trade-marks on the different slips vary in their likelihood of confusion, or in their deceptive similarity, or in the likelihood that the imitation trade-mark will be mistaken for the original. *This confusion may be due to the appearance of the trade-marks, their sound, their linguistic formation, their meaning or significance, or any combination of these four factors.* Neglect, however, all of these four factors except that of *sound*.

"Arrange the slips according to the likelihood of confusion in just the *sound* of the trade-marks: Put on the top of the pile the slip containing the imitation which shows the greatest likelihood of confusion in *sound* and on the bottom the slip containing the imitation which shows the least likelihood of confusion in sound. Between the top slip and the bottom put in order those slips containing imitations which show intervening degrees of likelihood of confusion in sound.

"In arranging the slips consider the trade-marks as if you have had no previous experience with them."

In addition, the observer was told that there was no time limit, and that he could take as long as he wished to make the arrangement, and further that he was allowed to rearrange the order until it satisfied him. To do the experiment usually required about five minutes. No information was given as to the purpose of the experiment.

Fifty observers assisted in the task of judging the trade-marks. They were mostly students of psychology at Columbia University, during the month of October, 1915. Some were undergraduates studying psychology, and some were graduate students with special training in it, and a few were assistants and instructors of psychology. Only four observers had not gone to college, but all had been acquainted with psychological experiments. No observer that took part in any of the recognition experiments acted again in this experiment.

2. *Results.* Table VIII gives the grades and probable errors of the word "Chero-Cola," each of the five infringing, and four non-infringing imitations. The first column in the table shows whether the imitation in the decision was held to be an infringement (I) or a noninfringement (N). The second column gives the original trade-mark in the decision; and

the third the imitative. The fourth column gives the grade; and the fifth and last the probable error of the grade. The marks are arranged in an order of probable confusion from most to least. The pair of trade-marks which received the smallest average was judged to have the greatest likelihood of confusion in sound; the pair which obtained the largest average was judged to have the least likelihood of confusion in sound.

TABLE VIII

THE GRADES AND PROBABLE ERRORS OF THE WORD "CHERO-COLA," EACH OF THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS

Decision	Trade-Mark Original	Trade-Mark Imitative	Av.	P. E.
I	Cascarets	Castorets	1.2	.03
I	Green River	Green Ribbon	3.3	.13
	Coca-Cola	*Chero-Cola*	*3.6*	*.17*
I	Listerine	Listogen	3.7	.15
N	Sozodont	Kalodont	5.1	.20
I	Gold Dust	Gold Drop	5.2	.12
N	Grape-Nuts	Grain-Hearts	7.0	.13
I	Club	Chancellor Club	7.8	.11
N	Mother's	Grand-Ma's	8.6	.09
N	Holeproof	Knotair	9.5	.08

"Cascarets—Castorets," an infringement, has been selected out of the 10 pairs of trade-marks as being the most confusing. It was placed in the first position by 42 observers, and in the second by 8. Its average grade or position is 1.2. "Green River—Green Ribbon" stands second with a grade of 3.3. "Coca-Cola—Chero-Cola" with a grade of 3.6 is nearly tied by an infringement, "Listerine—Listogen" with a grade of 3.7 for the third most confusing position. The grade of "Coca-Cola—Chero-Cola" is 0.3 lower than that of "Green River—Green Ribbon," and 0.1 higher than that of "Listerine—Listogen." These differences are not large. The grade of "Coca-Cola—Chero-Cola" is 2.4 lower than the grade of the most confusing infringement "Cascarets—Castorets;" and it is 4.2 higher than that of the least confusing infringement "Club—Chancellor Club." It is 1.5 higher than the most confusing non-infringement "Sozodont—Kalodont;" and it is 5.9 higher than the grade of the least confusing non-infringement "Holeproof—Knotair." The arrangement shows that "Coca-Cola—Chero-Cola" is about equally confusing with one infringement, less confusing than two infringements, and more confusing than two infringements and all four non-infringements.

A consideration of the probable errors will show that the grades of confusion of the different pairs of trade-marks have a high degree of validity. Moreover, the validity itself of the order can be measured. The probable error of the grade 1.2 is 0.03. It must, however, be remembered that the chances of the true grade being far outside this range decrease very rapidly. The possibility that the grade 1.2 of "Cascarets—Castorets" is due to chance is entirely negligible. Although all the remaining probable errors are somewhat larger, they indicate a high degree of validity of the average grades. The larger probable errors indicate that the order is less certain, and the difference in confusion between one pair of trade-marks and the next on the list is less. The probable error of the grade 3.6 of "Coca-Cola—Chero-Cola" is 0.17, or the chances are even that this grade is correct within about one-fifth of a unit.

The probable error of the difference 0.3 between the grade 3.6 of "Coca-Cola—Chero-Cola" and the next higher grade of "Green River—Green Ribbon" will show the chances that "Coca-Cola—Chero-Cola" deserves a grade as high as that of "Green River—Green Ribbon." The probable error of the difference 0.3 is 0.21. As the difference 0.3 is about 1.4 times greater than the probable error of the difference 0.21, the chances are about 3 to 2 that "Coca-Cola—Chero-Cola" does not deserve a grade as high as that of "Green River—Green Ribbon." As the difference 0.1 between the grade of "Coca-Cola—Chero-Cola" and the next lower grade of "Listerine—Listogen" is covered by the probable errors of the grades, the chances are nearly even that "Listerine—Listogen" deserves a grade as high as that of "Coca-Cola—Chero-Cola." The probable error of the difference 1.5 between the grade of "Coca-Cola—Chero-Cola" and the second lower grade of "Sozodont—Kalodont" is 0.26. As the difference 1.5 is about 5.8 times greater than the probable error of the difference 0.26, the chances are over 10,000 to 1 that "Coca-Cola—Chero-Cola" does not deserve a grade as low as that of "Sozodont—Kalodont." Therefore, the chances are about 3 to 2 that "Coca-Cola—Chero-Cola" does not deserve a grade as high as that of "Green River—Green Ribbon;" the chances are nearly even that "Listerine—Listogen" deserves a grade as high as that of "Coca-Cola—Chero-Cola;" and the chances are over 10,000 to 1 that "Coca-Cola—Chero-Cola" does not deserve a grade as low as that of "Sozodont—Kalodont."

Of the very well-known original trade-marks "Cascarets—

Castorets" alone stands higher than "Coca-Cola—Chero-Cola," and "Gold Dust—Gold Drop" and "Grape-Nuts—Grain-Hearts" are both lower. Of the less well-known original trade-marks "Green River—Green Ribbon" stands higher than "Coca-Cola—Chero-Cola," "Listerine—Listogen" about the same, and all the rest lower. In this relative position experiment the comparison between these two classes of trade-marks is not really necessary as the instructions required the observer to disregard the factor of familiarity with the trade-marks. In only three cases did the observer report any difficulty in so doing it.

TABLE IX

THE AVERAGE GRADES AND PROBABLE ERRORS OF THE WORD "CHERO-COLA," THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS

Number of Decisions	Decision	Av.	P. E.
	Coca-Cola—Chero-Cola	*3.6*	*.17*
5	Infringement	4.2	.12
4	Non-infringement	7.6	.11

Table IX gives the average grades of "Coca-Cola—Chero-Cola" of the five infringements, and four non-infringements. "Coca-Cola—Chero Cola" has a higher rank than both the average grades of the infringements, and the average grades of the non-infringements. It is 0.6 higher than the average of the infringements, and 4.0 higher than the average of the non-infringements. The average of the infringements is 4.2, and that of the non-infringements 7.6; their probable errors are 0.12 and 0.11 respectively.

The probable error of the difference 0.6 between the grade of "Coca-Cola—Chero-Cola" and the average of the five infringements is 0.21. As the difference 0.6 is about 2.9 times greater than the probable error of the difference 0.21, the chances are about 19 to 1 that "Coca-Cola—Chero-Cola" does not deserve a grade as low as the average of the five infringements. The probable error of the difference 4.0 between the grade of "Coca-Cola—Chero-Cola" and the average of the four non-infringements is 0.20. As the difference 4.0 is 20 times greater than the probable error of the 0.20, there are practically no chances that "Coca-Cola—Chero-Cola" deserves a grade as low as the average of the four non-infringements. "Coca-Cola—Chero-Cola," therefore, stands with high validity above the average of the five infringements, and the average of the four non-infringements.

In Table VIII, among the infringements the greatest difference is 6.6, between "Cascarets——Castorets" with a grade of 1.2 and "Club—Chancellor Club" with a grade of 7.8. Among the non-infringements the greatest difference is 4.4, between "Sozodont—Kalodont" with a grade of 5.1 and "Holeproof—Knotair" with a grade of 9.5. In Table IX the average of the five infringements is 3.4 higher than that of the four non-infringements. Thus the differences within either class of decisions are greater than between them. Two non-infringements, three if we include "Coca-Cola—Chero-Cola," are more confusing than the lowest infringement. The two most confusing imitations are infringements, and the two least confusing non-infringements. In sound confusion the results point consequently as did those in visual recognitive confusion to the unreliability and inconsistency of the legal decisions and of the present legal procedure. The difference, however, between the averages of the infringements and non-infringements in sound confusion by relative position is greater than in visual recognition confusion, and indicates a greater accuracy of the decisions.

A comparison of Tables III and VIII will show that the positions of the same pairs of marks in the two lists have changed in most cases. "Green River—Green Ribbon" is the only pair that did not change its position. "Coca-Cola—Chero-Cola" changed only one-half of a position, "Holeproof—Knotair" changed one position, "Listerine—Listogen" and "Gold Dust—Gold Drop" two positions, "Grape-Nuts—Grain-Hearts" and "Club—Chancellor Club" three, "Sozodont—Kalodont" four, "Mother's—Grand-Ma's" five and one-half, and "Cascarets—Castorets" six. The average of the number of positions changed for the 10 pairs of marks is 2.7. The lack of exact correspondence between the positions of the same pairs in both lists does not mean that the methods or results are contradictory or inconsistent. Though in both experiments presentation of material was through vision, each experiment principally affected two different senses. Exact correspondence would most likely have existed between the two lists if they were both concerned with measuring the same thing.

Conclusion

1. The word "Chero-Cola" shows a likelihood of visual confusion with the word "Coca-Cola." With one imitation in the experiment and when the mark is used without the name of the commodity, 28 per cent. of all the observers confuse in visual recognition the word "Chero-Cola" with "Coca-

Cola." With one imitation in the experiment and when the mark is used with the name of the commodity, 68 per cent. of all the observers confuse in visual recognition the words "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink." With 10 imitations in the experiments and when the mark is used with the name of the commodity, 40 per cent. of all the observers confuse in visual recognition the words "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink." Thus, either with or without the name of the commodity added to the trade-mark, with or without other imitations present, the word "Chero-Cola" shows a likelihood of confusion in visual recognition with "Coca-Cola." It should be observed that the per cent. of likelihood of confusion is not an absolute number representing the exact per cent. of individuals that would be confused in daily life. The per cent. holds approximately under the simple standardized conditions of the experiment.

2. Most of the observers confused are "absolutely certain" that they had seen the word "Chero-Cola," and only a few are either "reasonably certain" or have a "faint idea." That is, most of the observers put the greatest confidence in their wrong recognitions, and in so doing testify to or affirm their great belief that they are right.

3. Most of the observers confused state that they were "perfectly familiar" with the trade-mark "Coca-Cola" before they saw it in the presentation of the experiment, and only a few were either "moderately familiar" or "just familiar;" none were "unfamiliar" with it.

4. Confusion in visual recognition caused by "Chero-Cola" is greater than that caused by the imitations of the other three very well-known original marks; it is 12 per cent. greater than the confusion caused by the infringing imitation of "Cascarets," 14 per cent. greater than that of the infringing imitation of "Gold Dust," and 22 per cent. greater than that of the non-infringing imitation of "Grape-Nuts." "Chero-Cola" has a score of 8 per cent. less than the most confusing non-infringement, 6 per cent. less than the most confusing infringement, ties with another non-infringement, and has a higher score than four infringements and two non-infringements. The confusion in the case of "Coca-Cola—Chero-Cola" is about 7 per cent. greater than that of the average of the five infringing pairs of marks or of the four-non-infringing. The average of the infringements is only 0.7 per cent. higher than that of the non-infringements; practically this is no difference.

5. As determined by the average per cent. of observers that correctly recognized the duplicate and new marks, the task set the observer in the experiment is of moderate difficulty. In the three recognition experiments the average per cents. of corrct recognitions of the duplicate and new marks is 87, 79, and 85 per cents. The figures indicate then that the task set is not too arduous, nor exceedingly easy.

6. In measurement of confusion in sound by relative position "Coca-Cola—Chero-Cola" with a grade of 3.6 (P. E. 0.17) is in the third most confusing position in the list of 10 pairs of litigated trade-marks. It is ranked nearly equally confusing with one infringement, less confusing than two infringements, and more confusing than two infringements and all four non-infringements.

7. "Coca-Cola—Chero-Cola" has a higher rank of confusion than both the average of the grades of the infringements, and the average of the grades of the non-infringements. It is with high validity above the average 4.2 (P. E. 0.12) of the five infringements, and above the average 7.6 (P. E. 0.11) of the four non-infringements.

THE ANALYSIS OF TRADE ABILITY

By Edward S. Robinson, Yale University

I. *Introductory*

The control of learning and the measurement of acquired abilities are among the most important personnel problems before the industrial psychologist. He will find that these problems hold fully as great opportunities for the application of the method of his science as do the problems of intelligence testing. Although he has no ready-made formulas for dealing with the situation, he has a psychological, or better, an empirical point of view toward human behavior. By virtue of this point of view he can, if he will, contribute a systematic analysis of many problems of industrial behavior which seem so stubbornly to resist analysis on the part of the average man who is familiar with industry, but unfamiliar with psychological methods.

The possibility of applying psychological methods to the problems of acquired behavior in industry can best be illustrated by reference to the work of the late Trade Test Division of the Committee on Classification of Personnel in the Army. This organization had as its task the development of standardized tests for measuring the trade ability of men entering the National Army. It was the purpose of this work to increase the accuracy of selecting men best fitted for particular army jobs requiring special abilities. Civilian occupations seldom furnished exact counterparts of these army jobs, but correspondence was close enough and frequent enough to give a very high value to any information which could be obtained regarding a man's industrial experience. The trade tests were designed to check up on statements of soldiers concerning their pre-military occupations, and, further, to measure their proficiency in trades at which they claimed to have worked. The development of these tests required the observation, from a psychological point of view, of a great number of trades and tradesmen. I shall try to summarize here the results of those observations in so far as they apply, not to specific trades, but to trades in general.

Although it has some basis in a rather loose usage, the meaning of the term *trade* as used in this discussion is largely arbi-

trary. It does not, of course, refer to occupations mainly concerned with the exchange of goods, but rather to occupations emphasizing an acquired facility in the use of certain tools, instruments and machines for the achievement of certain classes of physical results. Moreover, this acquired facility or skill consists of a complex set of coordinations, intellectual, motor, or both, such as we find in carpentry, interior wiring and the like, and not of the very simple and oft repeated coordinations typical of the operation of much standardized factory machinery. *Trade ability* refers, then, to any complex set of coordinations which we find characterizing all men skilled in a given trade and marking them off as a single group.

II. *Defining Trades*

There are in existence scores of books and pamphlets which attempt to define the trades of modern industry, but scientifically most of these are of little or no value. Practically all of the writers of these definitions have fallen into an easy fallacy which appears when we look into their methods. Having a given title, say armature winder, before them, they have set out to collect from all possible sources descriptions of the jobs of armature winders. They have then tabulated these descriptions and issued them as a full and complete definition of the armature winder—as a guide for identifying the skilled armature winder wherever he may be found. But when we try to make such practical use of the definition, we find that it does not work. Only in negligible instances do we find the type of man we are looking for. According to the definition a skilled armature winder is a man capable of working on "A.C." and "D.C." windings, on "hand" and "form" windings, etc. Our natural inference from this is that the man who only partially fills these requirements is only partially skilled. We find, however, that there are men with all sorts of partial combinations of these enumerated requirements who are, nevertheless, classed as winders of the first rank by their bosses and co-workers. From this it is apparent that "armature winding" is not a single trade, but a name which has been applied indiscriminately to a variety of trades. The trade of armature winding, pure and simple, does not exist.

In order to frame useful trade definitions, then, the investigator must be thoroughly empirical. There is only one reliable source of information about the requirements of a trade and that source is to be found in the observed abilities

of individual tradesmen. Industrial experts of one kind or another may safely be consulted for a bird's-eye view of a trade whose definition is required, but such experts are not to be relied upon for detailed or exact opinions. The classification and description of workers according to their peculiar behavior equipments is a thing which must wait upon the field observations of men who have the training and patience to be inductive and empirical. Such a procedure will result in definitions or specifications quite different from those to be found in the large mass of current trade literature. We shall find that the tradesmen who can actually be identified in industry are "A.C." men, "D.C." men, form winders, hand winders, or general repairmen, etc. and not armature winders.

III. *Elements of Trade Ability*

The detailed work of tabulating the pertinent items of behavior of individual tradesmen requires a definite basis of analysis. In the Trade Test Division we found it convenient to conceive of ability in a trade as made up of a group of elementary abilities. These simple abilities, while not necessarily unanalyzable, are of such a nature that when further reduced they bear no apparent relationship to the complex trade ability of which they are parts. The riveting of the sheet metal worker is an element of his trade ability, although the behavior here involved is evidently open to further analysis. Such further analysis would, however, yield items of behavior with no apparent relationship to sheet metal working ability and, therefore, of irrelevant fineness. In all trades analysis stopped with the discovery of the simplest relevant units of ability.

From a psychological point of view the analysis of any trade brings out elements of two main sorts: (1) *Intellectual*, and (2) *Motor*. While in the strictest theoretical sense no sharp line can be drawn between these two categories, we may say roughly that the intellectual elements are the items of information and habits of judgment employed in meeting trade situations. Such items of information are displayed when a carpenter recognizes a mitre, a shingle hatchet, a valley rafter etc. Typical habits of judgment are displayed when an interior wireman selects one rather than another rosette as a ceiling fixture for a drop cord and when he decides to cover a bit of wire with porcelain tubing instead of lume. The motor elements are the coordinations involved in the unique muscular movements of tradesmen. The wood pattern maker possesses a large variety of nice coordinations which are

manifest in his manipulation of numerous chisels, planes and other tools. Shifting geers, accurate steering, etc., are motor elements of automobile driving.

When the Trade Test Division first faced the problem of measuring trade ability, a very confident opinion was expressed in all quarters as to the relative importance of intellectual and motor elements in trades in general. Psychologists, engineers, and tradesmen themselves insisted that the skilled tradesman is characterized by his manual rather than by his intellectual training, and that, on the other hand, there are large numbers of men having a thorough intellectual grasp of trades they cannot practice. The point seemed well taken and influenced appreciably the direction of our earlier attacks upon the problem of measurement. Concretely, this point of view appeared in an a priori confidence that performance tests would measure trade ability more accurately than oral or other tests emphasizing information and judgment. But increasing knowledge, gained from first hand observations of industrial conditions, soon made clear that this weighting of motor elements was based upon an outgrown state of affairs. Undoubtedly it was once true that ability in perhaps a majority of skilled trades was characterized by motor habits, to be acquired only after long and arduous training. This, however, is no longer the case. Pattern making, automobile driving and some other present day trades emphasize the motor, but far more common are the electrical trades, the machine trades, etc., which are largely intellectual and whose motor habits, with few exceptions, are relatively simple and easily acquired.

This statement that trades are becoming more intellectual may need reconciliation with the fact that modern industry calls for little intelligence from the run of her workers. Especially since the beginning of the war manufacturers have reduced the complexity of the duties of individual workers by cutting up trades into operations. Under such conditions it has been possible to use practically unskilled labor in the production of intricate mechanical products. But where jobs have remained complex and where workers have remained tradesmen in the sense in which I have defined that term, there has been a growing emphasis upon the intellectual as opposed to the motor. Each new machine introduced into industry tends further to emphasize the knowledge and judgment requirements of workers and to do away with skilled movements previously required.

IV. *Variations in Trade Ability*

In any trade ability ranges from that of the novice to that of the expert. The way in which trade ability develops can conveniently be studied by listing the elementary abilities manifest in men whose industrial status indicates just what stage they have reached in the learning process. In the novice we find no item of knowledge, no habit of judgment, no muscular movement which is a genuine element of the trade in question. The presence of such an element in a novice indicates one of two things: either the novice is not a novice, but an immature tradesman; or this particular element is extraneous to the trade and should, therefore, be excluded as a partial criterion of ability in that trade. In the expert we find a maximum number of genuine trade elements. In men whose status is something between novice and expert we find a partially complete group of trade elements.

The building up of a group of trade elements—the learning of a trade—takes place in a definite, fairly predictable way. The order in which the individual elements are acquired is not a matter of chance, but for the most part is the same for all tradesmen within a single trade. That is to say, most trade elements tend to appear at some definite stage in the learning process. Hardly any apprentice battery mechanic can do a successful job of "lead burning" on the top of a battery, while this same task is easy for almost any battery mechanic who has passed the apprentice stage. Similarly, technical knowledge of roofing appears somewhat suddenly in the process of learning house carpentry. These elements which tend to appear suddenly may be classified in the following fashion:

1. Trade Elements: Those possessed by practically all workers in a trade—apprentices, journeymen and experts—and by no one not in that trade.
2. Journeyman Elements: Those possessed by practically all journeymen and experts and by no apprentices or novices.
3. Expert Elements: Those possessed only by experts.

If for an element of any of these types we should plot the correlation between the probability of its appearance and trade status, we would obtain a sharply curved regression. Figures 1, 2 and 3 show the plotted correlations for hypothetical elements of each of the above types.

In our studies we found that the most frequent elements were those belonging to the first of the above classes. Elements tending to appear at the journeyman stage were next most frequent. Elements tending to appear at the expert

stage were quite rare. Increments of trade elements, then, grow smaller and smaller as progress is made toward expertness, and trade learning as measured by the acquisition of new elements is a negatively accelerated function.

Besides these elements, each of which has a definite point of appearance in the progress of learning, there are other elements whose probability of appearance, when correlated with trade status, gives a linear regression. Figure 4 shows a hypothetical element of this type. Such elements can be described as appearing in no novices, in a few apprentices, in more journeymen, and in still more experts. In our work we ran across comparatively few elements of this kind.

V. *The Measurement of Trade Ability*

Having once classified the elements of a given trade the construction of a test for that trade is simple. If the object of the test be that of separating men on either side of a point along the range of ability, it is necessary to include in that test only questions or job specifications calling for a display of elements tending to appear at just beyond that point. If, on the other hand, the object be that of separating men at several points along the range of ability, then the test should include elements tending to appear at just beyond each of these points, or elements whose probability of appearance increases gradually along the range of ability, or elements of both classes.

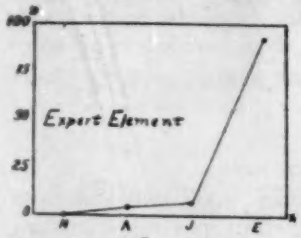


Fig. 3

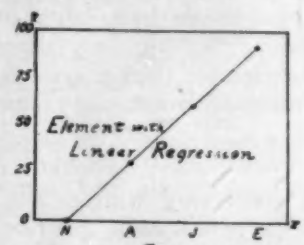


Fig. 4

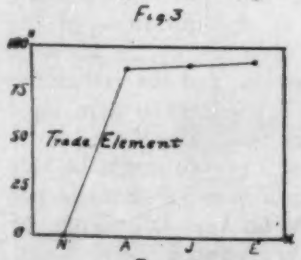


Fig. 1

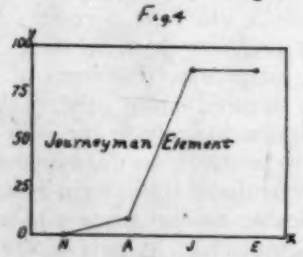


Fig. 2

x axes show stages in trade learning— N= novice, A= apprentice, J= journeyman, E= expert

y axes show probability of appearance in terms of percent

A WRITTEN TRADE TEST: MULTIPLE CHOICE METHOD

By J. Crosby Chapman, Western Reserve University; and Herbert A. Toops, Teachers College, Columbia University

The Army Trade Test Division was, on *a priori* grounds, extremely sceptical of written trade examinations. For this reason, in the theoretical examination, an oral test was used. The oral method has proved its power to distinguish, with a fairly high degree of certainty, between novices, apprentices, and journeymen, and to a less degree between journeymen and so-called experts. The possibilities and limitations of one type of written test form the subject of this article. In connection with the work of this Division an opportunity was given of trying out a written trade examination constructed on the multiple-choice plan. There are many advantages in being able to give group examinations of this kind. Among these are: (1) conservation of examiner's time, (2) reduction of scoring time and increase of scoring accuracy through the use of stencils, etc., (3) elimination of subjective judgments, (4) the possibility of rapidly eliminating the trade "bluffer," leaving only those who show evidence of some knowledge for further consideration or examination by means of oral, performance, or picture tests.

Assembly of Questions

In conference with expert tradesmen, sixty-two questions were prepared, each question having four possible answers, of which one was correct. Nearly all the questions of the army oral test for Bricklayers, partly collected by the authors, were adapted to this form of examination, and the remainder were secured from other sources. Trade terms were used almost exclusively in the four answers from which the choice was to be made, on the supposition that a person might be able to recognize a trade term from among a number of terms not applicable to the trade and yet have no real knowledge of its significance. A part of the test is reproduced.

BRICKLAYERS' TEST*

Directions:

In the sentences below, there are four choices for a correct sentence. Only one of the four choices is correct. Draw a line under the *one* choice which makes the best sentence. If you are not sure, *guess*; an omitted answer will count as a wrong answer in determining your score. The first two sentences, A and B, are answered correctly as a sample for you. Read them carefully, then go on to question 1, question 2, and so on to the end of the list.

Samples:

A. A half of a brick is called a: chunk, block, heel, bat.

B. Fire-bricks are laid in: concrete, cement, fire-clay, mortar.

Questions:

1. The top course of stone on a wall is called: coping, bond-stone, clip-course, capstone.
2. A brick set on end is called: upright, soldier, row-lock, stud.
3. Before plumbing up a corner, you should lay: 3-courses, 6-courses, 9-courses, 12-courses.
4. A brick that is set on the narrow edge is called: stretcher, oarlock, row-lock, header.
5. In coming to a height, if there is a course of brick difference in the level, you would call it a: haunch, filler, line-level, hog.
6. Filling-in the space between a front and back course of brick is called: slushing-up, tempering, tuck-pointing, plastering.
7. Bricks laid across the wall reaching to the face are called: stretchers, headers, layers, fillers.
8. A fire stop around a flue is formed by a: coping, skew-back, corble, indent.
9. If you have no metal ties in plain bond work, the front work of a building is tied-in with: king-course, clip-course, projection-course, belt.
10. In backing a 4-inch ashlar stone wall, you would tie it with: tie-joint, binder, brace, anchor.
11. A brick cut in half lengthwise is called: filler, soap, bat, chunk.
12. To keep the line level in the middle of a long wall you use: level, plumb-line, square, trigger.
13. In cutting a brick you would use a: chisel-bar, striker, brick-set, cutter.

---

*These eighteen questions were selected from the original set of sixty-two, which constituted the examination. This selection was made on the basis described in the text.

14.	The distance above the frame for a lintel should be:	¼-inch, ½-inch, 1-inch, 1½-inches.
15.	An ordinary stone sill should be bedded on the:	front, back, solid, two ends.
16.	To locate a building you use:	water-level, sighter, transit, line-level.
17.	A skutch is used in cutting:	cement, glazed-brick, stone, fire-brick.
18.	Another name for a blind header course is:	belt-course, clip-course, dental-course, hidden-course.

The answers to the above questions with the percentages of experts passing each are as follows:

1.	Coping—98%	7.	Headers—98%	13.	Brick-set—100%
2.	Soldier—98%	8.	Corble—85%	14.	½-inch—88 %
3.	3-courses—88%	9.	Clip-course—100%	15.	Two-ends—93%
4.	Rowlock—100%	10.	Anchor—85%	16.	Transit—88%
5.	Hog—100%	11.	Soap—85%	17.	Fire-brick—100%
6.	Slushing-up—98%	12.	Trigger—93%	18.	Clip-course—93%

The Subjects of the Experiment

The examination was given without time limit to 101 persons, distributed as follows: 35 novices without trade experience, 4 apprentices, and 62 trade union members. Of the novices, 29 were soldiers in an army training school, being instructed in other trades at the time, but inexperienced in the bricklayers' trade. These men were mechanically inclined, and above average in intelligence and education. It is obvious, therefore, that the questions were subjected to much more intelligent guessing than ordinary conditions, under which they would be employed, would furnish. Of the other six novices, 3 were stenographers who had several days experience, immediately preceding this test, in taking and transcribing stenographic notes relating to the general subject of bricklaying. The other 3 novices were clerical workers in the Department of Psychology of Carnegie Institute of Technology. These subjects were well informed concerning the technique employed in assembling such tests, and used various methods of "suggestion, inspection, and elimination" in efforts to make as high scores as possible.

Owing to the great scarcity of apprentices, only four could be examined. Their length of trade experience ranged from one month to five years. In view of experience and performance in the army oral test, the lowest apprentice can be con-

sidered little better than novice, while the best apprentice perhaps deserves to be ranked as a low journeyman.

The tradesmen, being all members of the union, more than fulfil the minimum entrance requirements in point of years of trade experience and skill. On the basis of their own written statements in regard to years of trade experience, type of work, and foremanship ability, these 62 men were divided by the authors into the two following groups: (a) 36 journeymen (hereafter called J.) or men of medium amount of skill and knowledge; (b) 26 experts (hereafter called E.), men of superior skill, knowledge and foremanship ability. This rather arbitrary classification was made in an effort to determine from the results of the examination whether a set of questions could be so selected as to differentiate not only non-tradesmen from tradesmen, and apprentices from skilled men, but also journeymen from journeymen experts, so divided.

Presentation of Results

Four points were allowed for each correct answer; there were no partial credits. Four points was selected as an arbitrary value because of army custom. A graph for each question was drawn, showing the average score made by novices, apprentices, journeymen and experts. Samples of such graphs are shown in fig. 1. Since little was known as to the most desirable combination of questions to make up the best set for an examination, five methods of combining varying numbers of questions were tried; only two of these methods are here described.

Method I is the entire set of 62 questions.

Method II is a selection, from the original 62, of 18 questions, chosen on the basis of the combined judgments of three persons experienced in similiar selection for the army oral trade tests. The requirements for a question in the army tests were: (1) increasing average scores from class to class, novices to experts; (2) a fairly low average novice score; (3) a comparatively high average expert score. The ideal type of question for the choice-method is shown by the straight line of fig. 1.

The distribution of scores by Method I are given in Graph I. By inspection, arbitrary lines of demarcation of N-A, and A-J may be made, corresponding to the *number of questions* correct as follows:

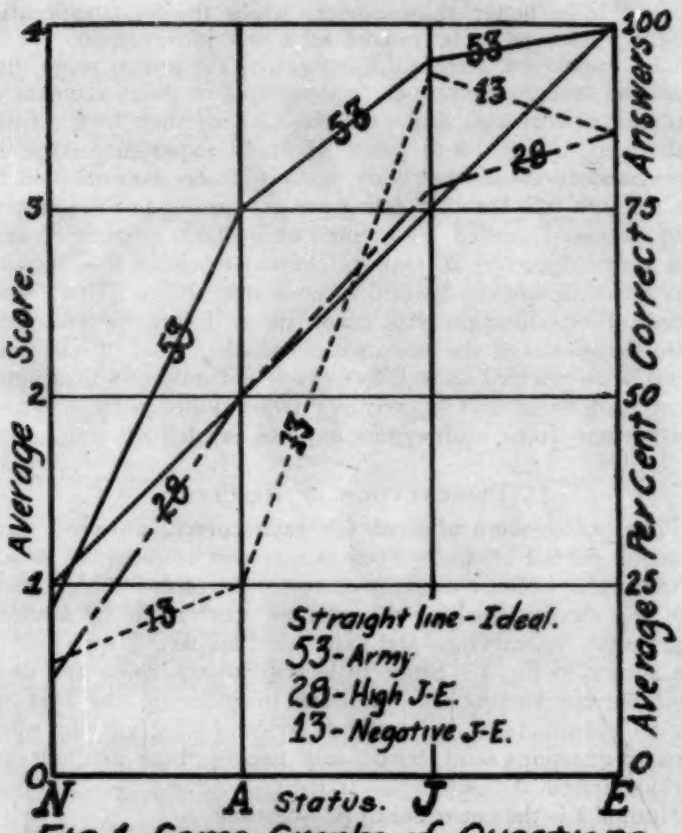
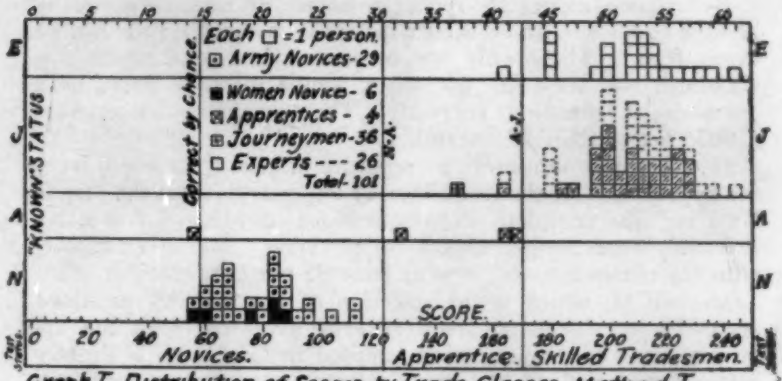


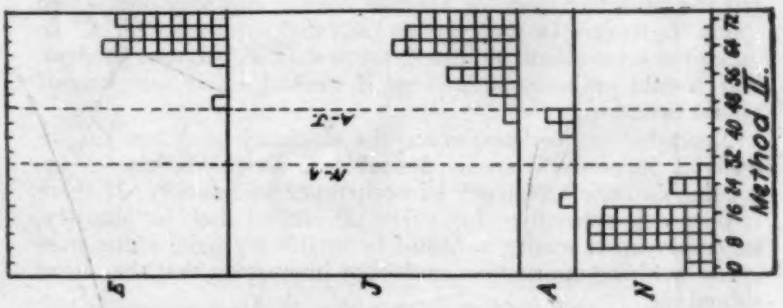
Fig. 1. Some Graphs of Questions.

Novice score	0-30
Apprentice score	31-42
Skilled tradesmen	43-62

It can be seen at once that the test does not differentiate between the journeymen and expert classes. Either the division into these two classes by judgments, based on the men's written statements, has nothing to do with the men's ability to answer this form of question, or the test as given is not sufficiently difficult to differentiate among skilled tradesmen. The homogeneity of the skilled tradesmen class as to scores can be easily seen from Graph I, where the expert distribu-



Graph I. Distribution of Scores, by Trade Classes. Method I.



tion has been added to the journeyman distribution by means of the dotted squares.

With the two arbitrary critical lines just mentioned, only three skilled union men out of 62, or less than 5 per cent, are displaced out of this larger class into the apprentice class. We can assume that all union men have had four years trade experience and possess a certain minimum requirement of ability. The apprentice, with only one month's experience, is here rated with the novices, as is only fair under the conditions. His guessing ability is as poor as the poorest novice, from which we can assume that he has learned next to no trade terms in one month of trade experience. Our own judgment is to the effect that intelligence is a better single measure of trade skill than are years of trade experience, after a certain minimum time for learning has elapsed.

By chance alone, the poorest novice should theoretically be able to answer 15 of the 62 questions correctly (one out of every four). Here, only two novices fail to come up to this standard, and they fail by only a single question each, both answering 14 questions correctly. The average novice answers a little better than 19 questions.

If the object is merely to separate skilled tradesmen from the apprentices, or the tradesmen of experience from the "trade bluffers," the complete set of questions of Method I is a remarkably accurate instrument for so doing. For this purpose, with the persons tested, it is at least 95 per cent efficient.

Method II, which is the selection of the best 18 questions according to the usual army criteria, gives us, with the arbitrary lines of demaracation adopted in Graph II, a slightly better separation of the three classes—novices, apprentices, and skilled union workmen. The one apprentice, with one month of trade experience, is still placed with the novices, but is now better than the average novice, and only one skilled union tradesman (a journeyman) is rated as an apprentice. To make the separations indicated, the test is 98 per cent efficient. This would probably be lowered if we had a fair sampling of apprentices.

Nevertheless, in determining the efficiency of a test for industrial purposes, it seems desirable to weigh the classes according to their frequency of occurrence in industry; if there is only one apprentice for every 25 skilled men in industry, an employment manager would be unable to make many mistakes by hiring apprentices under the impression that they were skilled men.

In industry, also, it seems desirable to measure the efficiency of test methods by taking into account the amount of damage caused by misplacement of a man by the test, some such measure as "dollar displacements," as suggested by R. W. Kelly. For it is a greater industrial mistake to give a journeyman or expert rating to an apprentice than it is to rate a novice as an apprentice; this may be seen from a consideration of the relative amounts of responsibility and wages involved.

From Graph II, Method II, we note that there is a slightly better differentiation of J. and E. than in Method I, but not sufficient to justify any arbitrary line of division between the two classes. If a line were drawn at 60-64, only two experts would be classed as J. while over half of the J. would be classed as E.

The final selection of questions, used in Method II and percentages of experts answering each correctly, are given in the

preceding sample. The average scores and average deviations of the trade classes in both sets of questions are shown in Table I.

TABLE I

AVERAGE SCORES AND AVERAGE DEVIATIONS THEREFROM, BY METHODS AND BY TRADE CLASSES

Trade Class	Method		No. of Cases
	I	II	
Novices	78.3 ±12.5	12.5 ±5.1	35
Apprentices	128.5 ±37.5	37.0 ±8.5	4
Journeymen	206.4 ±12.7	63.3 ±5.9	36
Experts	207.4 ±14.3	66.8 ±3.9	26
Possible Score	248	72	Total 101

SUMMARY

(1) A written trade examination, based on the multiple choice answer method, was given, without time limit, to 35 novices, 4 apprentices, and 62 journeymen.

(2) Two combinations of questions were tried—the complete set of 62 and a specially selected series of 18.

(3) Both combinations, with a very high degree of accuracy, separate the novices, apprentices and journeymen, but neither of them differentiate between the journeymen and the so-called experts, although the eighteen-question set is slightly more efficient.

(4) In the absence of extreme language difficulty, the written group examination can be used with great advantage to (1) eliminate the bluffer, (2) to separate the apprentice of average attainment from the journeymen.

The authors acknowledge the help accorded them by Messrs. T. J. Kirby, R. Pintner, and R. M. Reinhold, all members of the Army Trade Test Division.

MINOR STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

III. The Relation of the General Intelligence of School Children to the Occupation of their Fathers

By S. L. Pressey and Ruth Ralston [1]

1. *Problem and Materials.* It has been generally recognized by workers with tests of intelligence—from Binet on—that results with such tests from school children may vary surprisingly from one neighborhood to another; children in one neighborhood may average as much as three years in "mental age" above children in another type of neighborhood.[2] It has usually been argued that such differences were but further evidence of the superior native ability of the people living in the more prosperous communities, and of the inherent incapacity of the poor. That is to say, a direct correlation has been thought to exist between the economic and sociological status of a neighborhood or locality and the intelligence of its people,— as measured in its school children.

Such community comparisons, by means of mental tests, are as a matter of fact, decidedly interesting, and have value both psychologically and sociologically. The test findings may, apparently, reflect local conditions with surprising detail.[3] It is obvious, however, that to select children by neighborhoods is to select them very roughly indeed; if we wish to study the relation between the mental endowment of school children and the economic and sociological status of their families we should select our cases according to some economic or social standard which can be applied to each particular child or family in question. Children might thus be grouped

[1] Help in the collecting of the data was also given by Mr. J. Ward Starr.

[2] See for instance W. Stern, *The Psychological Methods of Testing General Intelligence*, pp. 50-57; Terman, *The Measurement of Intelligence*, pp. 114-118; Yerkes, Bridges and Hardwick, *A Point Scale for Measuring General Intelligence*, pp. 73-83.

[3] See for instance, Pressey, S. L., A Comparison of Two Cities and their School Systems by Means of a Group Scale of Intelligence, *Journal of Educational Administration and Supervision*, Vol. V, 1919, pp. 53-62.

according to the income of the family, according to home conditions, according to the occupation of the wage earner, or with reference to any similar criterion.

Results from such a study should contribute not a little to our understanding of local differences in the general intelligence of school children. The findings should, however, have an interest quite as great in their bearing on other sociological problems. If, for example, we measured the intelligence of the adults of a community, we would expect (if the present social unrest has any adequate basis) to find not a few men and women in positions much below the level of their abilities, and incapable individuals not infrequently in responsible places. Test measurements of the exactness desired can hardly be obtained directly from adults, because of the specialization of function that the workers already on the job have acquired, and because of the tendency of mental tests, thus far, to favor those in the more clerical lines of work. But, if mental ability is hereditary, then measurements of the *children* of these workers should give the information desired. If, for example, the brightest child in a hundred should be found to be the child of a day laborer such a discovery would surely be of decided interest from many points of view.

The present paper thus attempts: (1) to obtain some measure of the average differences in general intelligence found between children coming from different kinds of homes, and (2) to examine the amount of over-lapping between the different groups — the extent to which children of every degree of ability may appear in each type of family.

The scale of intelligence used in this study was developed at Indiana University during the past year.[4] It will be described in detail in connection with the analysis by test given below. The city surveyed is a place of about 12,000 inhabitants. The business of the town centers around a large furniture factory, but there are also quarries near, and the place is the county seat and the center of trade for the county. The State University is also located in the town. There is thus presented a variety of occupations and of economic levels that could hardly be found in the study of the same number of children in a single section of a large city. It should also be added that the town contains almost no foreign element; the troublesome

[4] See the report of the Seventh Annual Conference on Educational Measurements, Indiana University, April, 1919, or Cross-out Tests: With Suggestions as to a Group Scale of the Emotions, *Journal of Applied Psychology*, Vol. III, June, 1919, pp. 138-150.

factor of language handicap is to be considered almost negligible.

The survey was made in April 1919. All the children from the third grade through High School were tested — a total of some 14,000 cases. In the study of the present problem, however, no children above fourteen or under ten have been considered, since shortly after becoming fourteen (the legal age limit) the children in the poorer families frequently leave school; the children remaining are thus a selected group. A somewhat analogous difficulty prevented the inclusion of the younger children, eight and nine years old. Many of these were still in the first two grades where the scale could not be used; and since the factors being studied might well show themselves in school retardation, it was felt that the younger children tested were again a selected group, as far as this particular problem was concerned. There were a total of 707 children within these age limits (10-14 inclusive) in the city schools, examined with these tests on the day of the survey. The occupation of the father of each one of these children was looked up in the school record, the records being checked by one of the writers, who is a teacher in the system, and well acquainted with many of the families. A total of 548 cases, the occupation of whose father was known, were found.[5] These 548 cases were now classified in four groups according as the father was (1) a professional man, (teacher, lawyer, doctor, minister, editor), (2) an executive (independent business man, foreman), (3) an artisan (electrician, engineer, skilled workman), or (4) laborer (section hand, factory operator, unskilled laborer).

2. *Results.* The children were distributed by age and occupational groupings as follows:

[5] The desired information could not be secured for the remaining 159 children within the five ages selected for study; they had moved away since the tests were given and the cards had been withdrawn, or the card failed to show the occupation of the parent—or the data was otherwise incomplete. The writers can see no reason, however, for supposing that these omissions have affected the trend of the results in any important particular. The occupation of the parent was chosen as the basis of classification from the various possibilities mentioned above largely because it was definite and because data could be obtained in the largest proportion of the cases. No other single fact, of this importance for the problem, could have been secured for even half the number of cases it was desired to include in the data.

Age	10	11	12	13	14	Total
Number	125	116	107	99	101	548
Occupation	Professional	Executive	Artisan	Laborer		Total
Number	57	105	138	248[6]		548

In handling the results the total score made on the examination has first been considered, and has been taken, for the moment, to represent a final rating of "general intelligence."

From a table showing the distribution of scores at each age, for each group, the medians, ten and ninety percentiles were calculated for the entire number of children at each age. The per cent of children in each group scoring above the median for their age was then worked out as shown in Table I.

TABLE I

PER CENT OF CHILDREN IN EACH OCCUPATIONAL GROUP ABOVE THE MEDIAN FOR THE ENTIRE GROUP

	Ages					
Occupational Groups	10	11	12	13	14	All Ages Combined
Professional	93	85	80	81	86	85
Executive	58	44	56	76	68	68
Artisan	52	32	48	38	40	41
Laborer	38	41	39	34	40	39

It will be seen that the results are what is to be expected as far as the average of general mental ability is concerned. The children of professional parentage test as superior while the children of day laborers test as relatively poor in mental endowment. Apparently, there is a direct correlation between occupation and mental ability.

But the more important question as to the compactness of the distributions and the amount of overlapping from class to class is still to be considered. The per cents of children in each occupational group who make scores in the highest and lowest ten per cent for their age run as follows:

	Prof.	Exec.	Art.	Labor.
In highest 10%	41	10	8	6
In lowest 10%	1	2	12	17

[6] The large plurality of the unskilled laborers would in itself seem of some significance in connection with the second problem above. A multitude of factors are undoubtedly involved. But there are surely suggestions here that more men are in the lowest type of ocupations than we would expect, from a normal distribution of ability to be capable of no more than the least skilled work.

The results of this analysis are most interesting. There was one child of a professional man who scored in the lowest 10% for his age; otherwise the children from professional homes form the bulk of the highest ten per cent. The more compact distribution of the scores in this group is due, perhaps, to the fact that the professions have more definite requirements than do the other occupations as regards general intelligence. At the other extreme in almost every respect is the group of children whose fathers are unskilled laborers. These children give scores ranging from the very poorest to the best; there are some children of common laborers who score in the highest ten per cent for their age. These facts would seem to confirm the statement, made frequently, that in the present day industrial world, the level of work is much below the level of ability of the worker. It would certainly suggest that many individuals were not where they belonged, — were employed in positions calling for much less than their best.[7]

However, with such striking results, one naturally questions the validity of the tests, and wonders whether the nature of the scale may not have influenced the findings in some illicit way. The four tests composing the scale were selected after extended study as being highly differential of mental subnormality and superiority, and relatively free from the influence of other factors. Still it is conceivable that the tests might reflect home environment to a greater extent than is usually thought. Analysis by test should help not a little in answering this question.

The four tests of the scale may be briefly described. The first consists of twenty-five sentences with the words disarranged, and among the words of each sentence one word that cannot be used in the sentence (So "boy was sky the sick"). The subjects are told to cross out this extra word. The second test is an information test; it consists of twenty-five lists of five words each. Of the five things named in each list four are alike in some important way and one thing is

[7] The smaller proportional representation of the laboring class in the fourteen year group is of interest in this connection; children of unskilled laborers make up 50% of the twelve year group, but only 37% of the fourteen year group. And though it is especially the duller children who thus drop out this appears to be not altogether the case— there is a thinning of the whole distribution. That is, the children of laboring people who are quite as bright as children from more favored homes have much less educational opportunity. A similar analysis including children above fourteen would be interesting in this connection.

different; this one is to be crossed out (so "dog, cow, horse, oak, cat"). The third is a test of arithmetical ability. It consists of twenty-five series of digits, the digits being arranged according to some rule, but there being in each series one number that breaks the rule; this number is to be crossed out (so "2, 4, 6, 7, 8"). The fourth is a test of vocabulary and of moral discrimination. It consists of twenty-five lists each of five moral terms (as "begging, buying, trading, cheating, borrowing"); the subjects are to cross out the worst thing in each list.

It might be objected, now, that a good or a poor score on such a scale might be quite as much the result of a good or a bad home environment as of a good or bad mental endowment. Particularly the first and the last tests might be sensitive to such environmental factors; they would appear to be, to a considerable extent, reading and vocabulary tests, and a prosperous home offers greater opportunities for reading and the acquirement of general culture. Test II, however, should be much freer from such influences, as there are no words used in this test that are not in the vocabulary of the average child; in fact, it is conceivable that the child from a poor home, who knocks about the streets, might be able to pass the test rather more successfully than the more sheltered child from a good home. And it is difficult to see how the third test could be helped by a good environment to any appreciable degree. If, then, the tests are measuring environment as well as innate ability the first and last tests should show greater differences between the occupational groups than is shown by the two middle tests. But if the difference shown by each one of the tests is about the same then it would seem fair to conclude that the factor causing this difference was of a general and fundamental character, presumably mental endowment.

Table III shows the per cent of children from each one of the four occupational groups who score, on each test, above the age median for the whole group. The five ages have been combined for the sake of simplicity.

TABLE III

SHOWING THE PER CENT. OF CHILDREN FROM EACH OCCUPATIONAL GROUP WHO SCORE ABOVE THE MEDIAN FOR THEIR AGE (FOR THE TOTAL GROUP) ON EACH TEST

Occupational Group	Test I	Test II	Test III	Test IV
Professional	80	80	80	87
Executive	58	64	68	56
Artisan	56	54	51	59
Laborer	47	46	48	45

As may be seen from this table, the results are highly consistent from test to test. The results by test would suggest, therefore, that the scale really was measuring some fundamental underlying factor such as native endowment, and that any special features of environment that might be expected to influence the record made on certain of the tests have not affected the scores to any important extent. The children in the first group (professional) are superior on every test and the children of the last group (laborer) are inferior on every test; the other groups also hold their place remarkably well.

3. *Discussion.* Such an analysis by test seems to give something of a measure of the "reliability" of a scale for such purposes; in so far as the findings are consistent from test to test of those tests used, in just so far similar findings might be expected if other tests still were employed.[8] Too much must not, of course, be made of such inferences. It may very well be that subtle matters of attitude, differences in the total fund of information, and other influences immediately resulting from differences in home environment, play no small part in such findings. But such differences are, after all, so fundamental as to be of hardly less importance than differences in native intelligence.

The writers are convinced, however, that such environmental influences can also be specifically studied, with profit, by means of "group tests." An informational test to measure the extent to which the more expensive conveniences, and luxuries, are known in the home would seem not at all impossible,—or a test to measure the amount and type of reading done outside of school. Tests of very little children, as they first enter school, should also get much closer to such differences in the intimate home influences.[9] The possibilities of the group test as a general method of investigation have not, the writers feel, been adequately appreciated. A multitude of problems, psychological, educational, sociological, await investigation by such means.

4. *Summary.* The paper reports an analysis, according to the occupation of the father, of results with a group scale of intelligence from 548 unselected school children 10-14 years

[8] That is, unless all the tests used were measuring some one special ability. A series of intercorrelations between the tests used in the present study will shortly be presented, to show that such is not true of this scale.

[9] See the article referred to above, in the June number of this JOURNAL.

of age. All of the children of these ages for whom complete data were available in a city of 12,000 population were included. It was found that:

1. Of the children of professional men, 85% made scores above the median for the total group; of the children of executives (independent business men, foremen) 68% rated above the median for the total group; of the children of artisans (unskilled workers) 41% rated above the median, and of the children of laborers only 39% rated above the median.

2. The groups overlapped largely, though not completely. Both extremes of the distribution contained some members from each occupational group: however, the artisan and laborer groups predominate at the lower end of the distributions while the professional and executive groups show a majority of the high scores.

3. Analysis by test shows these findings to be constant for all four tests. It is argued from this that the findings have some general reliability.

IV. A Practical Information Test

For Use with Delinquents and Illiterate Adults

By S. L. Pressey and I. M. Shively*

1. *Problem and Materials.* Some three years ago the writer, while engaged in psychological work at the Boston Psychopathic Hospital, became much interested in tests of information to be used with delinquents. He came to feel that the Terman-Childs vocabulary test [1] was distinctly unfair in such cases. For such cases (and perhaps for the average healthy out-of-door child) it was not, he felt, an unselected list. A dictionary contains a greater proportion of literary words and of technical terms than is found in ordinary talk, especially among the lower classes; a dictionary covers primarily a reading, rather than a speaking vocabulary. An "unselected" list from a dictionary will, then, in proportion as it is unselected, show this same specialization. The result is a distinct unfairness to the gang boy from the slums or the girl off the street; they not infrequently possess a bulk of information—unfortunate and otherwise. But they do not know Terman's list. The rating on his scale suffers accordingly, and the examiner obtains a distinctly false impression as to the individual's fund of ideas. The present paper is a brief description of an attempt to formulate a test, of the general nature of Whipple's Range of Information test,[2] which should be especially applicable to such cases.

The list was, in fact, expected to be of value for three purposes: (1) It should serve as a test of general intelligence, especially applicable to children from poor homes and to adults of scanty education. (2) It should, if the words were properly chosen and grouped so as to cover important "fields" of

* The writer (Dr. Pressey) is indebted to Miss Shively for the scoring and tabulating of the results—in fact, for all the real labor of the study—the work being done in connection with a summer school course in *Mental Measurement.*

[1] See Terman, L. M., *The Measurement of Intelligence.* Houghton Mifflin Co., 1916, pp. 224-231.

[2] Whipple, G. M. *Manual of Mental and Physical Tests,* 1915 Vol. 2, pp. 317-322.

information, be of distinct value in the investigation of an individual's interests, and in sounding the extent to which he is acquainted with various types of experience — good or bad. (3) In the third place, it was hoped that the test might yield data of some value in connection with efforts to differentiate deteriorated individuals in work with the insane. Roughly it may be said that mental deterioration shows itself in a fund of knowledge much greater than would be accounted for by present capacity for acquirement of ideas. Such cases should, then, be distinguished by a relatively good record on tests of information along with very poor records on tests of learning. As an informational test to be used for this purpose such a list as has been just suggested should be more satisfactory for the general run of psychotic cases than Terman's more literary list. The intention was to investigate the value of the test for each one of these three purposes. It has not been possible for the writer to complete the study as he had planned because of the many pressing demands of work in other directions. But a description of the test and the tentative results thus far obtained may be of some interest.

The list of words is shown below. It was arrived at by a study as to which fields of information were most familiar to the general run of cases at the Psychopathic hospital, by in-

RANGE OF PRACTICAL INFORMATION

Moral	Religious	Governmental	Legal	Mechanical
Stealing	Creed	Mayor	Jury	Pulley
Lying	Disciple	Senate	Plaintiff	Plow
Gambling	Benediction	Constitution	Probation	Scythe
Cheating	Heretic	Tax	Larceny	Rudder
Pity	Sacrament	Tariff	Sheriff	Piston
Cruelty	Altar	Naturalization	Perjury	Forge
Revenge	Monastery	Consul	Embezzlement	Dynamo
Envy	Bishop	Regent	Blackmail	Insulation
Patriotism	Easter	Quorum	Forgery	Derrick
Duty	Trinity	Treaty	Bail	Semaphore

Industrial	Commercial	Recreational	Domestic	Public
Piece-work	Bank	Ace	Oven	Post-card
Floor-walker	Check	Dice	Yeast	Telegram
Corporation	Credit	Short-stop	Flue	Ambulance
Line-man	Pawn	Touch-down	Sirloin	Steerage
Broker	Mortgage	Jockey	Borax	Editor
Strike-breaker	Auction	Cueball	Thimble	Insurance
Boycott	Contract	Odds	Dimity	Vaudeville
Stoker	Bankruptcy	Lager	Mop	Taxi
Foreman	Discount	Stogy	Tape	Pension
Rake-off	Lease	Rag-time	Eaves	Militia

quiry of doctors, social workers — and also patients — there, and by newspaper reading. It is by no means as satisfactory as might be desired, and a double meaning for several of the words has been discovered; but it seemed unwise to try to improve it further without trial in definite form. The writer at first wavered between two methods in selecting the words. It was possible (1) to try and obtain words, within each field of information, which any intelligent person of fair general information should know. Or, (2) the list might be made up of words which only an individual especially conversant with that field would be able to define. If the list were made up in this last way, and used either as an information or a free association test, it might well have some detective value—might bring to light a hidden interest in unfortunate subjects, perhaps. With the formation of such a list in view the writer for some weeks carried a special pad of paper on which he jotted down all the highly special words which he could obtain from interviews with various unsavory characters appearing at the hospital, and with street-car men, clerks, sailors, mechanics, and others having their own special supply of odd words. But it was finally decided that the first concept was the better: the second basis of selection would lead to a list of very limited usefulness.

2. *Results.* As a test of general ability and a check on the results of Terman's vocabulary test (the use for which the list was first intended) the value of the list would best be demonstrated by data from children of the poorer classes or from adults of little schooling; such cases might be expected to do very poorly with the Terman words, but to give results with the writer's list much more in keeping with their mental age. The writer has been unable to obtain any systematic data of this sort, but has instead tried to investigate (a) the value of the list as it may be used with average school children, and (b) its usefulness as a means of studying differences in interest and special information. In investigating the first problem the Terman list of words, and the 100 words of the above list, were given as a group test to 56 children in the fifth grade in two schools in Bloomington, Indiana. It was desired to find out (1) the most satisfactory method of scoring, (2) the correlation of results from the two lists with some independent measure of general ability, and (3) the extent of the specific contribution made by each list to the correlation.

Two methods of scoring the results were used. In the first place, they were scored according to Terman's rule, one point being allowed for each word which was defined with a degree

of exactness sufficient to indicate some knowledge as to its meaning. That is, scoring was very lenient and no partial scores were allowed. The results were then gone over a second time, a score of 1 being given for a bare indication of acquaintance with the word, and two points being allowed when the word was correctly defined. The correlation of the results with the practical information test, scored in these two ways, was .86. The correlation for the Terman list was .84. These correlations, of course, suggest strongly that it makes little difference which method of scoring is used.

Results were now correlated with findings from a group scale of general intelligence given to these same children a few months previous.[3] The correlation of the practical information test, scored in the first way mentioned, with these ratings was .40. The correlation, scored in the second way, was .45. The results would suggest that the more exact method of scoring might be expected to give more significant results. The Terman list scored in this last way was now correlated with the scale. The correlation was .50. Apparently for school children (if we may generalize from these results) Terman's list is slightly better as a test of general intelligence; a result which surprised the writer, but of course, does not affect the question as to the relative merit of the two lists for the special purpose already mentioned.[4] A higher correlation in both instances would be expected from older children, since results from older children would be less affected by irrelevant cir-

[3] Pressey, S. L. and L. W. A Group Point Scale of General Intelligence. *Journal of Applied Psychology*, Vol. II, 1918, pp. 250-269.

[4] Whatever the comparative merit of the two tests for any of these purposes, the Practical Information test is at least vastly superior in the number of extraordinary ideas it brings to light. In the present experimentation the Terman list resulted in only one mild epigram: hysterics is "sick in your head." But with the writer's list such gems abound. To pawn is to "sell things temporarily;" credit is "to pay for something some other time;" a mortgage is "when another person can take what he wants off you." Larceny is "stealing things that don't amount to much:" cheating is "to take money from people who can't count good." The religious terms bring out vividly the prevailing cynicism among the younger generation: an altar is "an object on which sacrifices were made in old times—in use now;" and Easter "is the day on which Christ *was supposed to have risen!*" Two curious bits of etymology also appear here: a disciple is "when it is not coupled together," and a bishop is "the owner of a shop." Yeast is "put into bread to make it grow." A floor-walker "quiets the baby;" rag-time is "bouncy music." An editor "puts the news into the newspapers." A boycott is "a boy that has a uniform on and will be a soldier some day;" patriotism is "to be kind, and put out a garden." The finding of such choice bits gives a real zest to the labor of scoring.

cumstances such as differences in ease of writing. But since the children were allowed all the time they needed there is no reason to suppose that any special difficulty in writing or in expression played any very serious part in the results.

A mild effort at partial correlation resulted in a correlation of .78 between the two lists, $r_{13 \cdot 2}$ was found to be .26 and $r_{12 \cdot 3}$ was .11, the first variable being the group scale, the second being the practical information test, and the third the Terman vocabulary. Apparently the practical information test does make its own contribution to the correlation. Correlation with an individual examination such as the Binet, and oral definition for both lists, would, of course, be much more valuable. However, the above figures will give some rough idea as to what might be expected in general of the comparative correlation of each informational test with ratings of general ability.

An effort was next made to discover whether clear differences in knowledge of special fields might be found by means of the practical information test. Analysis by sex was first tried. In making this analysis data from the 56 fifth grade children already mentioned was combined with results from 29 eighth grade children. The median for the boys in each group of ten words, for each grade, was found, the number of girls above this median counted, and the total number of girls above the median expressed as a per cent of all the girls tested. The results were as follows:

Word Group—	1	2	3	4	5	6	7	8	9	10
Per cent girls above boys median	50	50	25	17	20	25	50	20	61	58

There are here clean-cut sex differences. The girls are distinctly superior in knowledge of the "domestic group" of words, and slightly superior in the list having to do with public utilities and information of that same general nature. Boys are superior in knowledge of the "governmental" words, of the legal and mechanical words, and also of those having to do with recreation.

A comparison of the fifth and eighth grade was also tried. But it was impossible to find any definite difference in increase of knowledge along different lines. The children in a small colored school were next given the list, a total of 36 children in all. Of these 14 were in the fifth grade, 15 in the sixth, 6 in the seventh and 1 in the eighth; the group was, therefore, lumped and compared with the results from the fifth grades. The per cents of colored children scoring above the median

for the white children, in the ten groups of words, were as follows:

	1	2	3	4	5	6	7	8	9	10
Per Cent:	45	56	81	56	19	80	81	81	72	80

To the writer, at least, the results are distinctly surprising. The colored children average about 1 grade above the white chidren and over two years older. Nevertheless, they show a poorer average knowledge of moral terms, about the same acquaintance with religious terms. They show what might almost be called a total lack of acquaintance with the general vocabulary of mechanics. With the small number of cases these results should, of course, not be made too much of; it should be remembered also that the words are graded in only the most rough fashion and the test otherwise loosely organized. But it may at least be said that the results are suggestive. The writer cannot but believe that the races differ in the tendency of their interests; and he feels decidedly that a valuable way to investigate such a matter is to measure the information acquired along various lines.[5]

3. *Discussion.* A word remains to be said on these few findings. They surely suggest fundamental differences, as regards information and interest, between the sexes and between the two races. The writer feels strongly that until we know more about differences in interests we cannot know what we should with regard to sex and race differences. The tendency is to investigate such matters by somewhat isoteric and roundabout methods. We use the free association test, for example, act mysterious, employ a stop-watch, and otherwise embarrass the subject. A straight-forward investigation of information might very well, in such instances, bring the same data with no difficulties of cooperation. As one method of investigation the writer would like to suggest such lists as those above. In its present form the test here described is

[5] It should be stated that in making these analyses the second method of scoring was used, that is, a child might get a zero, one, or two, on each word. As indicative of the general fall of the figures it may be added that the median for the fifth grade, when the practical information test was thus scored, was 42.5; for the eighth grade it was 106 (out of a total possible of 200 points). The fifth grade medians for the separate groups were as follows:

	1	2	3	4	5	6	7	8	9	10
	12	3	2	1	5	0	4	0	8	5
For the Eighth:	18	11	12	5	11	6	12	9	14	17

The fifth grade was composed of 37 girls and 29 boys; the eighth grade group of 14 girls and 16 boys.

obviously not satisfactory; it is not adequately organized, it is hard to score and awkward to handle; but it seems to have very good possibilities as a method of investigation. It may have a practical value too. But the writer has been diverted from that matter by the research interest of the data. He hopes to have further material to offer later.

4. *Summary.* The paper may be briefly summarized;

1. It is argued that an "unselected" vocabulary test is unfair to delinquents and to adult cases of little education. As a substitute a selected vocabulary covering certain fields of common information is offered.

2. A comparison of this test with Terman's vocabulary test is made. It is found that fifth grade children give results, correlating more highly with findings from a group scale of intelligence, with the Terman vocabulary test than with the writer's information test.

3. Distinct sex and race differences in the various fields of information are found.

4. It is suggested that such a test offers a convenient method for the investigation of sex and racial differences in interest.

V. The "Efficiency" of a Group Scale of Intelligence in Prognosticating Success and Failure in Junior High School

By S. L. Pressey

1. *Problem and Materials.* The paper deals with the prognostic value of a group scale of intelligence previously described in this *Journal,*[1] for forecasting success or failure in Junior High School. This scale was given in March, 1918, to all the children in the sixth grade (the entering grade) of a large Junior High school — a total of 165 cases. In the fall of 1919 the subsequent school history of these cases was investigated. Of the original number 20 cases were found to be lost to the study, because the family of the child had moved away, or for other similar reasons having no connection with school standing (cases leaving school because of school failure have been included in the study, as will be described below). For each one of the remaining 145 children the school marks for June, 1918, January (semesters) 1919, and June 1919 were obtained. The paper has to do with the relation between these marks and the test scores.

Two questions have been put: (1) What test scores were made by those children who failed in school (that is, using the school's own definition, those children who obtained marks below 75 in three or more of the four "promotion subjects" —arithmetic, geography, history, language)? (2) What is the relation of test scores to later success in school as shown by the last available marks (June, 1919), or when this class as a whole had completed the seventh grade? In each instance the effort has been to present the findings in terms that would be readily interpretable for use in individual diagnosis; in fact the interest of the writer has been rather in these methods than in the particular results dealt with.

2. *Results:* (*a*) *Prognosis of School Failure.* For the purpose of studying the relation between test score and failure in school the distribution of test scores for the 145 cases was divided into ten parts by calculating the deciles throughout.

[1] Pressey, S. L. and L. W. A Group Point Scale for Measuring General Intelligence, *Journal of Applied Psychology,* Vol. II, 1918, pp. 269-287.

The number of failures scoring in each decile was then determined. The results run as below. As has already been said, the tests were given in March, 1918. The first row below gives the number and distribution of failures for June, 1918; the second row gives the same facts for January, 1919, and the third row the data for June, 1919. The last row summarizes by giving the position of the scores of all the children who failed during any one of these three terms.

	Decile Division										Total
	1	2	3	4	5	6	7	8	9	10	
Failed:											
June 1918	9	6	3	1	2	0	1	1	0	0	23
Jan. 1919	6	4	3	1	1	0	0	1	2	0	18
June 1919	9	6	2	4	1	0	0	2	1	1	26
All failures	11	9	5	4	2	0	1	2	2	1	37

It should be added that this final row does not count any child twice, but shows the number failing at any one or all of these three terms. It should also be said that the 1919 rows include as failures those who failed in June, 1918 and left school because of the failure.

Perhaps the most striking thing about these results is the bimodal distribution of the failures. The vast majority of the children who fail tested well at the bottom of the distribution. But there is another small group near the top. There is a suggestion here of a maladjustment not only between dull child and school, but between school and superior child as well — perhaps some of the brighter children do not find sufficient stimulus to their abilities in the routine school work of their "normal grade." Anyhow, this little group of failures who test as above average in ability is distinctly interesting.

The matter is, however, somewhat aside from the problem in hand. The practical problem is this: What indication can be obtained, if such an examination be given to children in their first year in Junior High School, as to those pupils who will later fail in their school work? For such a practical diagnostic purpose the writer would like to suggest the following table, which is simply the last row of the table above, expressed in terms of per cent of the number of cases in each percentile division who were failures.

	Decile Division									
	1	2	3	4	5	6	7	8	9	10
Per cent. failing, or—per cent probability that a given case will fail	76	62	35	28	14	0	7	14	14	7

Thus there were 11 failures in the first decile division; there were a total of 14.5 cases in each decile division,—and 11 is 76% of 14.5. The other per cents were similarly found. The table shows the incidence of failure throughout the entire distribution of test scores. It is also, however, readily interpretable for purposes of individual diagnosis. Suppose a principal gives the scale to the entering class in his Junior High School during their first year. He may assume (if the school from which the writer's data is taken may be considered reasonably representative) that if a given child scores below the ten percentile for the grade there is a 76% probability that that child will, within the next two years, fail at least in one semester of work. If another child scores between the thirtieth and fortieth percentile there is a 28% probability that that child will fail — and so on for the other decile divisions. The method is direct and obvious, and would seem of some general usefulness.[2]

(*b*) *Prognosis of School Attainment as Measured by School Marks.* In order to obtain some more general indication as to the relation of test scores to future school accomplishment the relation of these scores to the marks obtained by all the children in June, 1919 — almost a year and a half later —was studied. The median mark of the marks obtained, at this time, by each child, in the five subjects arithmetic, reading, grammar, history, and geography, was used to represent school success. And a correlation was run between these median marks and test scores. The correlation was found to be .50.

However,—as has been well pointed out recently[3]—a correlation coefficient is a very inadequate way of presenting such data for purposes of individual diagnosis. In order to obtain a statement more satisfactory, for such a purpose, the scatter diagram from which the correlation was plotted was divided into five parts each way, by ruling lines through the 20, 40, 60, and 80 percentiles of each distribution. And the table was made to read as for 100 cases, by expressing the number in each compartment of the diagram as a per cent of 145—the total number of cases. The result was as follows: (I being best and V poorest in each array):

[2] A somewhat different application, and to a psychiatric problem, has been made in a previous article, "Irregularity as a Measure of Deterioration" *Jour. of Abnormal Psychology*, December, 1918.

[3] Thurstone, L. L. "Mental Tests for College Entrance," *Jour. of Educ. Psychol.*, Vol. X, 1919, pp. 129-42.

DIVISION INTO FIFTHS—TEST SCORES

		V	IV	III	II	I
Division into fifths— school marks	I	0	0	2	5	13
	II	1	4	4	8	3
	III	4	4	5	4	3
	IV	7	6	5	1	1
	V	8	6	4	2	0

If, now, the hypothetical principal above mentioned has given the scale to his sixth grade, and finds that a particular "problem case" scores in the upper fifth of the class, he may conclude that the probability is 13 out of 20 that that child will, at the end of the next school year, be receiving marks which place him in the upper fifth of his class. There is only 1 chance in 20 that this child will be marked below the middle group. There are 0 chances out of 20 that the pupil will appear in the poorest fifth, in school grades,—and so on.[4] The method is direct, readily understood by teacher or principal; and not only the probable diagnosis, but also something as to the definiteness of the indication, is shown.

As with the previous method, a larger number of cases is necessary to make the table reasonably reliable—the ideal form would probably involve decile divisions, the table then reading as for 1000 cases, or 100 in each array. But the writer feels that the methods themselves should be of distinct practical value, particularly in dealing with the very concrete and definite prognostications of business and vocational psychology.

3. *Discussion.* A word remains to be said with regard to such use of school marks, and school failure, as criteria in terms of which to evaluate tests of "general intelligence." It might be argued that the peculiar value of such tests was in in their ability to get beyond the particular narrow set of facts by which school marks are usually determined. And the ultimate criterion becomes—what the child *would* do, in a system highly adaptable to individual differences. But we also may wish to know what children, *things being as they are,* will fail next year, a year later; we may wish to know the children now in the eighth grade who will fail in high school

[4] It should be said that eight cases leaving school because of failure in June, 1918 have not been included in the correlation. Strictly speaking, they should be, somehow, probably at the bottom of the distribution of marks; our hypothetical principal does not know in advance who is going to leave school and should be provided with a prognosis table, covering *all* the cases under consideration But for present purposes the omission of these eight cases may be considered negligible.

should they go there, the college entrants who will fail their Freshman work. It may very well be that for prognosis over a brief period, in a situation such as is dealt with in the present paper, a set of achievement tests would give a more definite indication than tests of more general content; the writer hopes soon to present data bearing on this point. The study of such very concrete and rigidly defined problems is certainly much needed; such study would appear the only empirical and scientific way of determining the validity of the concept of "general intelligence"—that is, the validity of the theory that prognosis problems can be generalized.

4. *Summary.* 1. The paper deals with scores obtained from the sixth grade children of a large Junior High school, and with the school history of these children for three semesters since the tests were given. A definite statement of the prognostic value of the tests, over this period, is sought.

2. A numerical statement is obtained of the probability of failure in school implied by scores at various points on the distribution of test scores.

3. An analogous statement for prognosis of marks, over one year, is presented.

BOOK REVIEWS

JAMES BURT MINER. *Deficiency and Delinquency, An Interpretation of Mental Testing.* Warwick and York, Baltimore, 1918, p. 352.

The author here presents a comprehensive study of mental testing as applied to deficients and delinquents, assembling from a vast body of literature the various conclusions thereto appertaining. As the title suggests, the study is an interpretative one. We may add that it is also a wholesome critique. His claim to greater practicality in this important field of psychology is verified throughout, as there is shown to be a place for the serious attempts of social welfare workers, teachers, and amateurs, as well as the trained clinician.

Heretofore, the great complexity of the problems of deficiency has not been adequately solved by the employment of objective or qualitative tests alone, since the results have always been vitiated by certain errors necessarily connected with examinations. While the earlier objective tests have contributed much to basic data, and while they are sufficiently accurate to serve as a check to empirical judgment respecting obvious cases of deficiency, they are relatively worthless for the diagnosis of that larger group of borderlines, the accurate testing of which alone determines the true value of a test. The central theme and purpose of the study is, therefore, to set forth and explain the requisites of a system of tests which will serve to diagnose the troublesome "borderline" cases. The need of redefining the borderlines of deficiency becomes obvious to the author through practical experience with deficients in the Minneapolis Juvenile Court. Valuable data is also gathered from the Glen Lake Farm School for juvenile delinquents. As a further background for this study, the author has had the advantage of visiting scores of institutions in Austria, England, Switzerland, Germany, France, and Italy. To all this may be added the author's knowledge gained through the study of law, and the legal problems now fast becoming coextensive with social problems.

Accepting the common conception of feeble-mindedness, which holds it to be a lack of development rendering the individual socially deficient and requiring constant supervision or guidance, and excluding from this category the so-called "conative" cases of deficiency, the author's contribution may be summarized as follows:

1. Deficiency and delinquency are not mutually inclusive. Deficiency may be regarded as a state of permanent undevelopment, while delinquency may or may not accompany it. In order to eradicate subjective opinion in the differentiation of the two, a conservative borderline on the basis of Binet or other tests, should first be established, so that treatment of the delinquent is assured a still further individual consideration.

2. Present-day tests are inadequate for reliable diagnostic purposes, since they tend to exaggerate qualitative differences among abnormal subjects. Deficiency may now properly be looked upon as a relative and "quantitative" condition; every degree and shade of undevelopment is the rule.

3. The number of defectives in proportion to the school and general populations have been variously estimated; and these estimates indicate

something of the illusory and unscientific nature of tests and their interpretation. Therefore a "*percentage definition* of tested deficiency is suggested for determining the borderline below which an individual may be presumed to be so deficient as to justify isolation." The most conservative estimate of the number of defectives may now be said to be: "Those who would at 15 years of age be in the lowest 0.5% in tested ability among a randomly selected group, may be presumed to be so deficient as to justify isolation. Above them the next 1.0% may be regarded as uncertain, since the bulk of them would require some supervision or guardianship during life." The verification of such borderlines may be regarded as the main contribution yielded from the new data here studied.

4. Practically speaking, the location of borderlines on a test scale is symptomatic only, since different communities offer different results.

5. Emphasis should be placed on the social significance of deficiency both in its diagnosis and ultimate results. Deficiency and delinquency should therefore be referred to a commission consisting of expert medical and psychological assistants, so that conative cases may not be unjustly retained in public institutions permanently. To this end, a "temporary home school" is advisable.

6. Legal definitions of the criminality of deficients should be so changed as to obviate all subtle questions of mental analysis, such as: "was laboring under such a defect of reason as not to know the nature and quality of the act he was doing or know the nature of the wrong" (New York Law). Such law places too much emphasis upon the intellectuality of the accused, whereas in most cases of delinquency, the crime is more closely related to the emotions. The problem of criminality is a "problem in the treatment of hereditary diathesis in which the mental deficiency is the largest factor."

7. Whether society should adopt a policy of sterilization, or that of isolation, by force or consent, is an open question. At present it appears to be a question which may well await the perfecting of developmental tests that may assure accurate diagnosis. But wherever the cases are obviously permanently deficient, isolation is advisable for girls and women of child-bearing age.

8. External influences, such as criminality in the home, character and amount of education, standards of living, etc., seem to be less important than is heredity.

9. The schools have in all cases of deficiency, a special work to do in the discovery of such pupils, and in the referring of the same to the proper authorities. The school may also do a great deal in gathering data relative to the various correlations of conduct and mental attainment, which so far have been relatively unscientific.

University of Vermont. J. W. Sprowls.

Sherwin Cody. *Commercial Tests and How to Use Them.* School Efficiency Monographs. World Book Co., 1919.

The author presents valuable tests for determining ability to perform fundamental operations in business offices, viz., tabulation, reproduction and following of instructions, elementary invoicing, elementary arithmetic, elementary English, business letter writing, stenography, typewriting, penmanship and filing. These tests are in line with the tendency to test specific abilities and have the further qualification of being short. They have been standardized by use among thousands of grammar school and high school graduates and office employees of several large business houses, the average performances in the several

tests being given for each group. However, in giving these norms the author fails to add any measure of variability, thus detracting materially from the scientific and practical value of his work.

The author's use of the expression "business ability" to describe these tests is possibly misleading. They do not test business *capacity* but rather achievement and are therefore classifiable with educational or proficiency tests. Used alone they have rather limited value for the prospective employer, since they give no indication of an individual's future value as an employee nor of his ultimate vocational fitness for a business position. General intelligence, learning ability, habits of industry, etc., must be obtained from other sources or be assumed to correlate highly with the scores in the tests. However, improvement rate in a series of tests, together with school marks and teachers' ratings, could be used successfully by employers for selecting employees who have a high degree of capacity.

One test in the series, the Tabulation Test, is called an 'intelligence' or 'mental alertness' test. This amounts to stretching the conception of intelligence tests rather severely, but the statement is in line with the author's contention that a "high degree of accuracy in figures, ..., is an evidence of mental capacity," and shows "a discipline in mental control which must carry over into all the activities of life." Even if this statement is accepted it cannot be assumed that the reverse is equally true, for high mental capacity and intelligence is possible and frequent where ability to use figures with speed and accuracy is lacking.

An outstanding feature of the book is the method used for grading and checking the tests in the schools. The pupils themselves do the checking. This makes for speed and, in the author's opinion, gives a sufficiently high degree of accuracy. For example in New York, in 1917, "the tests were given to 6,000 eight-grade pupils, three tests to each. The margin of points overlooked or wrongly marked was not sufficient to change the median for any class as much as one per cent."

A gross error appears in the form of directions for giving the tests (page 57): "Tell pupils they must be as careful not to get too high a mark as not too get too low a mark, for then on the next test they will not show the improvement which is equally important with a high grade to start with."[1]

Carnegie Institute of Technology, Pittsburgh, Pa. H. C. Kenagy.

William Fretz Kemble. *Choosing Employees by Mental and Physical Tests.* The Engineering Magazine Company, New York, 1917, p. xiii+333.

The purpose and claim of the book is perhaps best expressed in the author's own words: "The following pages reproduce a series of tests which not only will pick out the man of great capacity from among a mixed company undergoing the test, but will arrange a staff of men in the approximate position of their ability when certain weights are assigned according to the nature of their work." This claim seems totally unsupported by results. Kemble not only does not prove in any way that his tests will do what he claims for them if handled by an experienced man; but he gives no accurate and sufficiently detailed instructions; therefore if the tests were given by an inexperienced man, there would be no hope of their receiving a proper evaluation even were the tests adequate, which we believe they are not.

[1] Review approved by the Committee of the American Psychological Association on Publications in Applied Psychology.

For example, Kemble nowhere states how he has standardized his tests. In several cases he gives the "human average" but says nothing of how this average was obtained. He suggests a psychogram of the subject, but does not indicate how this psychogram can be evaluated by the business man. And what is equally important, he does not seem to recognize any limitations to tests, nor does he differentiate between the precision of different tests, being willing to give an exact score in spelling ability and an equally exact one in honesty.

In order to so estimate a man's ability, Kemble would use four sources of information. The inadequacy of these is apparent to any one who understands the laws of testing. They follow:

I. An employment blank, consisting of five parts:
   a) a personal history blank
   b) a financial statement
   c) an expenditure blank
   d) a questionnaire concerning the religious and moral life of the person
   e) a personal belief blank

II. The "Peg Board Tests."

III. A series of mental activity tests. There are twenty-one of these tests, most of them of the ordinary type: memory for sense material, etc. However, a few are of rather an exceptional nature. The general information test consists of a series of questions, the information for answering which would be most apt to be picked up among day laborers—for example: "Give the proportions used in making mortar." The ethical test has questions similar to the following: "If you were about to take a trolley and a friend got off a car and offered you a transfer, would you use it?" The personal opinion test is rated in an original way; the answers to the questions are decided by the vote of at least ten high-class executives. The test then is supposed to differentiate between the executive and the popular type of mind. One's honesty is rated by the number of the following questions that he answers by "Yes": "Do you smoke? Do you drink? Do you use slang? Do you swear?" Kemble's method of using these tests is to decide what elementary qualities are involved in any occupation for which he is testing applicants, pick the tests which test these qualities and sum them for the applicant's score for that occupation. Anyone at all familiar with testing realizes at once the danger of such a proceeding.

IV. A study of the physical characteristics of the subject, type of face, length of fingers, etc.

In general, Kemble has made a rather unique collection of tests. A few have new features which perhaps offer ideas, but his claim of 90 per cent accuracy and the lack of standardization makes the book of little value for practical purposes, and even pernicious if used by an inexperienced person.[1]

Carnegie Institute of Technology, Pittsburgh, Pa. M. A. Bills.

Frederick Schleiter. *Religion and Culture, A Critical Survey of Methods of Approach to Religious Phenomena.* Columbia University Press, New York, 1919, p. 206.

The subtitle of the book should be borne in mind when one is reading it. It is the methods of approach to religious phenomena that the author is discussing. And the first method taken up is that propounded

[1] Review approved by the Committee of the American Psychological Association on Publications in Applied Psychology.

by Buckle, the studying of limited areas intensively and then drawing general conclusions from the results obtained. This method "bristles with fallacies and insupportable presuppositions, the most obvious and far reaching of which embodies the idea that the ethnographic phenomena found in some specially selected area, are the result of an independent development, and constitute, as it were, an indigenous entity possessing complete historic individuality" (p. 10).

The comparative method which consists "in bunching of static facts to indicate the common properties of primitive mind," . . . "has served to bring together similar customs, rites and ideas the world over, to stimulate investigation of them and has, in many cases, laid the basis for their elucidation" (pp. 21, 27). But it should be remembered that all genealogical hypotheses "lack objective confirmation and that their relative merit is to be judged largely on the ground of obvious plausibility and ingenuity" (p. 40).

The framing of ethnographic analoga the author characterizes as "a somewhat romantic procedure" (p. 67).

Animism and the mana theorists are equally incompetent to solve the problem, since the former "is far from literally true" (p. 80), and the latter "are full of inconsistencies, obscurities, and ambiguities" (p. 123). The discussions of animism and mana are very interesting and instructive.

The rôle that the concept of causality has played in religion and culture is truly great. This very obscure and ambiguous concept has lately come in for a great deal of criticism, and some scholars have gone so far as to suggest that it be rejected altogether. Mach says: "There is no cause nor effect in nature; nature simply has but an individual existence; nature simply is" (p. 142). Russell, who is perhaps one of the most severe critics, points out that "in advanced sciences such as gravitational astronomy, the word 'cause' never occurs" and that the "concept is not employed at all nor does it find embodiment or asylum in the algebraic equation" (p. 144). "Fiske characterized the development of science as a progressive 'de-anthropomorphization' and Comte supposed that the evolution of society involves the continuous elimination of 'fetishistic' elements" (p. 146).

Lastly the author discusses "the convergence of mental processes which lead to the attribution or predication of a dynamic relation between two or more elements" (p. 153). In this connection he calls attention to Dresslar's work on superstition among persons "professionally trained for the work of teaching." This work shows that 44.9 per cent believed either wholly or partially in certain superstitions. But "neither the character of the connecting link nor the manner of the achievement of the results appears in the tables" (p. 157). In another connection the author points out that "in many cases, he does not know precisely how these events, objects or processes are connected,—in other words, the nexus does not rise into his consciousness" (p. 161).

But there is not only an unconscious acceptance of the articulating mechanism, there is also to be found "a highly conscious and rationalistic enquiry in which the mind attempts to grasp, by acts of deliberate apprehensions, the causes of an event and then launches out boldly upon a more or less boundless path" (p. 164). Several interesting illustrations are here produced, both among primitive men and men of our own time, all of which go to show that an "enormous range and diversity of mental processes are involved in causation" (p. 191).

The purpose of the book is to "suggest that these logical processes,

which lift facts out of their indigenous habitats and deal with them in their universal aspects, should be applied with more critical caution than is customary" (p. 193). The positive contribution of the book is naturally not as great as the negative. But it constitutes a good critical survey of the field it covers and is therefore a valuable contribution to the literature on this subject.

KARL J. KARLSON.

The following publications have been received:[1]

JAMES WINFRED BRIDGES. *An Outline of Abnormal Psychology.* R. G. Adams and Company, Columbus, O., 1919.

W. J. CRAWFORD. *Experiments in Psychical Science, Levitation, Contact, and the Direct Voice.* E. P. Dutton and Company, New York, 1919.

L. M. GILBRETH. *The Psychology of Management.* Sturgis & Walton Company, New York, 1918. Reprinted without changes from 1914.

HENRY C. LINK. *Employment Psychology.* The MacMillan Company, New York, 1919.

NORAH H. MARCH. *Towards Racial Health.* E. P. Dutton & Co., New York, 1919.

DANIEL STARCH. *Educational Psychology.* The MacMillan Company, New York, 1919.

LEWIS M. TERMAN. *The Intelligence of School Children.* Houghton Mifflin Company, Boston, 1919.

MARGARET J. HAMILTON. *Psychological Analysis and Re-education, with Case Studies.* Reprinted from *The Journal of Educational Psychology,* Feb., 1919.

M. LAIGNAL-LAVASTINE. *The Internal Secretions and the Nervous System.* Trsl. by F. T. Robeson. Reprinted from *Nervous and Mental Disease Monograph Series.* No. 30, 1919.

S. D. PORTEUS. *Cephalometry of Feeble-Minded.* Reprinted from the Vineland, N. J., *Training School Bulletin,* June, 1919.

——— ———. *Porteus Tests—The Vineland Revision.* Reprinted from the Vineland, N. J., *Training School Bulletin,* No. 16, September, 1919.

ARTHUR H. SUTHERLAND. *Ungraded Rooms in Los Angeles City Schools.* First Annual Report, Division of Psychology, Year ending June 30, 1919.

J. E. WALLACE WALLIN. *Meeting the Needs of the Mentally Handicapped Child in School.* Reprinted from *Ohio Bulletin of Charities and Correction,* June, 1919.

---

[1] Mention here does not preclude further comment

———— ————. *The Field of the Clinical Psychologist and the Kind of Training Needed by the Psychological Examiner.* Reprinted from *School and Society,* April 19, 1919.

———— ————. *The Achievement of Mental Defectives in Standardized Educational Tests.* Reprinted from *School and Society,* Aug. 29, 1919.

J. HAROLD WILLIAMS. *The Intelligence of the Delinquent Boy.* Reprinted from *Journal of Delinquency,* Monograph No. 1, 1919.

CARRY ON, A Magazine on the Reconstruction of Disabled Soldiers and Sailors, Vol. 1, No. 10, July, 1919 (probably the last isssue).

REVISTA DE PSIQUIATRIA. Lima, Peru, Vol. 2, No. 1, July, 1919.

REVISTA DE EDUCACION. Santo Domingo, R. D. Vol. 1, Nos. 3 and 4, 1919.

NOTES

A special committee of the National Research Council, consisting of Dr. R. M. Yerkes, chairman, and Dr. M. E. Haggerty of the University of Minnesota, Dr. L. M. Terman of Stanford University, Dr. E. L. Thorndike of Teachers College, Columbia University, and Dr. G. M. Whipple of the University of Michigan, with financial support from the General Education Board, have formulated a plan for using the army mental tests in schools. Such intelligence tests have been used in schools for some time on individual children, but the new plan provides for handling them in groups, even whole class-rooms at a time. The committee selected about twenty tests for careful trial. This trial was made on five thousand children. As a result the committee has now been able to select from the tests two series which seem to be the most satisfactory and these will now be tried on several thousand more children in order that they may be further perfected before they are finally offered to the teachers of the country for general use.

This carefully worked out program for group tests will make it possible and practicable to make wholesale surveys of schools annually, or even semi-annually, so that grade classification and individual educational treatment can be adjusted with desirable frequency. It is expected that the methods will be ready to be published for general use early in 1920. The army tests on which these new group tests for children are based and which were used with striking success and advantage during the war, were originally devised by a group of psychologists working under the auspices of the National Research Council.

The following statements are extracted from a report by Vernon Kellogg of Leland Stanford Jr. University published in *The Nation's Business,* November, 1919.

"The lesson of the war as regards science, then, is first, that the efforts of scientific men can be greatly speeded up by a proper stimulus; second, that these efforts can be made immensely more effective and fruitful by a proper organization; and, third, that while such an organization can be encouraged and helped by the Government, it can be effected by cooperative effort among scientific men themselves.

"The founding of the National Research Council is the outgrowth of this lesson. And the special characteristic—and a truly American one—which distinguishes it from the other rather similar organizations of England, Canada, Australia and Japan, and also from the already long-established government scientific bureaus of our own country is that, although officially recognized by the Government, it was not initiated or organized by it and is not supported by it. It is the outcome of a nation-wide cooperative effort of the scientific men of America, including representatives of the universities, the various special privately endowed research institutions, the scientific laboratories of the great industries and the Government scientific bureaus, and the numerous unattached specialists. And it is entirely controlled, in a thoroughly democratic way, by these many cooperating scientific

investigators. It derives its support from funds contributed from private sources.

The Council is organized primarily as a permanently constituted part of the National Academy of Science.

"Effective prosecution of the Council's work requires not only the cordial collaboration of scientific and technical men of the universities and special research institutions all over the country, but also those connected with the scientific and technical branches of the Government.

"As now organized, the National Research Council has permanent headquarters in Washington, with an executive staff of scientific men giving their whole time to the work of their respective positions. The Council's field of activities is divided among thirteen divisions, gathered into two main groups; first, one of six "general relations" divisions, and second, one of seven divisions devoted to special lines of science and technology. Each of these divisions has a resident chairman and a small office staff in Washington and a number of non-resident members.

"There is a Division of Anthropology and Psychology, as yet without a chairman,* but with a list of members chosen from among the leading psychologists and anthropologists of the country."

Dr. W. V. Bingham, Head of the Division of Applied Psychology of the Carnegie Institute of Technology at Pittsburgh, has been appointed Chairman of the Division of Anthropology and Psychology of the National Research Council. Dr. Bingham is an authority on methods for measuring the intelligence of normal adults. Early in the war, as Secretary of the Committee on Classification of Personnel of the Army, Dr. Bingham applied these methods to testing the mental capacity and fitness of recruits as a basis for assignment and training for particular military duties, and later continued this work as Lieutenant-Colonel in the Personnel Branch of the General Staff.

At present Dr. Bingham is connected with several bureaus of the Carnegie Institute, which are engaged in studying the application of these principles in commercial and industrial occupations. One of these bureaus, that of Personnnel Research, is supported by annual contributions from 30 corporations. This bureau is engaged in ascertaining the best methods for selecting and developing executives, salesmen, and clerks.

Another bureau is applying the same principles in developing methods for selecting and thoroughly training workers in about 900 positions in seven of the leading department stores of Pittsburgh. These stores contribute $32,000 annually for these investigations. The financial support given this work of investigating and applying scientific mental tests reveals the confidence which is being placed by corporations and store executives in the money value of a rational study of their employment problems.

An appropriation of fifty thousand dollars in five annual payments has been granted to the Iowa Child Welfare Research Station of the State University of Iowa, of which Dr. Bird T. Baldwin is director, by the National Women's Christian Temperance Union, for the purpose of making detailed scientific investigation into the development and conservation of child life. The gift is for research and the station will

* Since this statement was first published, such appointment has been made. See notice following.

be given a wide range of freedom in the selection of its problems, which will center around the field of eugenics, nutrition, and the mental and social development of the child. Dean Carl E. Seashore is chairman of the advisory committee.

The Society for American Fellowships in French Universities offers each year for open competition among graduates of American colleges and other suitably qualified candidates a number of fellowships, not to exceed twenty-five, for the purpose of encouraging advanced study and research in French Universities.

The fellowships, of the annual value of $1000, are granted for one year and are renewable for a second year. They may be awarded in the following fields of study:

Anthropology, Archaeology and History of Art, Astronomy, Botany and Agriculture, Chemistry, Criminology, Education, Engineering, Geography, Geology, History, Law, Mathematics, Medicine, Philology: Classical, Romance, Oriental, Semitic, English,—Philosophy, Physics,—Political Science: Economics and International Law,—Psychology, Religion, Sociology, Zoology.

Fellows will be required to sail to France not later than July 1st of the year in which the award is made, to matriculate in a French University for the following session, and to pursue studies in the field of science designated in their awards. They will be expected to send accounts of their studies together with reports of their progress from their instructors.

Applicants, men or women, must be citizens of the United States and at least twenty years of age. They must be:

1. Graduates of a college requiring four years of study for a degree, based on fourteen units of high school work; or,
2. Graduates of a professional school requiring three years of study for a degree; or,
3. If not qualified in either of these ways, must be twenty-four years of age and have spent five years in an industrial establishment in work requiring technical skill.

Applicants must be of good moral character and intellectual ability, and must have a practical ability to use French books.

Applications must be made on Application Blanks furnished by the Society and must be accompanied by:

1. A Certificate of birth;
2. A Certificate of Naturalization, if needed;
3. A Certificate of College Studies;
4. A Certificate of Industrial Work, if needed;
5. Testimonials to Moral Character and Intellectual Ability;
6. A Photograph, signed and taken within a year; and
7. Printed or written articles, theses and books, written or published by the applicant.

Applications should reach the Secretary of the Society not later than January 1st of the year in which the award is to be made.

Application Blanks and further information about the fellowships may be obtained from the Secretary,

Dr. I. L. Kandel,
576 Fifth Avenue, New York.

The Massachusetts Society for Mental Hygiene, of which Professor William H. Burnham of Clark University is the president, held a Conference on Mental Hygiene in Worcester, Mass., November 14th, 1919. Dr. Walter S. Fernald, Superintendent of the Massachusetts School for the Feeble-minded, outlined a state program for the care of these unfortunates which involves a statistical survey and official record of all feeble-minded in the state, supervision and care of those who are not inmates of institutions, with the help of social and charity organizations, provision for special training during the formative period, and the inauguration of local mental clinics. Dr. Burnham showed that success and failure in school were important conditions influencing mental health and pointed out that present day administration and methods presented too many opportunities for failure and did not sufficiently utilize the mental stimulus of success. Dr. E. E. Southard, Director of the Massachusetts State Psychiatric Institute in Boston, summarized and classified the most important causes of mental ills under the headings of disease, ignorance, vice, legal entanglements, and poverty, which most frequently operate in pairs but only rarely alone, and showed that successful attack upon one member of a pair usually results in a great diminution, if not entire disappearance, of the other member. Dr. A. W. Stearns concluded the program by a brief summary of the aims and methods of the Society.

It is a matter of common knowledge that the increased cost of production of to-day is in large measure due to the demand on the part of labor for shorter working hours and more pay. The common way of the manufacturer of meeting this increased cost has been to pass it all and a little more on to the consumer by greatly raising the price of his commodity.

At the Athenaeum Press, where Ginn and Company's books are manufactured, there has just been inaugurated a plan of training foremen, forewomen, and other factory supervisors in production methods. A three-months' course is being conducted by the Business Training Corporation of New York for a group of forty-six, composed of the treasurer, press manager, department heads, and all foremen from the man in charge of a room of fifty to the one who supervises the work of two or three. The work consists in the study of especially prepared subjects, the solution of problems relating to each subject, and the discussion of this material at six bi-weekly meetings. At each meeting a lecture is delivered by an experienced production man. Afterward the meeting is thrown open to discussion of the material set forth in the text, and problems which may be peculiar to any particular branch of work in the plant. The subjects covered are teamwork, handling men, machinery and materials, organization, production records and management.

Beside a more thorough all-round knowledge of production methods, the course gives an opportunity for the leaders in the Press to get together, discuss important problems, not only among themselves, but with the higher executives, and to acquire at first hand the spirit of unity and team play which in the last analysis means the elimination of unrest.

INDEX OF SUBJECTS

(The subject headings of all contributions except Book Reviews and Notes are printed in SMALL CAPITALS.)

ABSOLUTE INTELLIGENCE SCALE 118
Accuracy, Test for Mental 179
ADULTS, AVERAGE MENTAL AGE OF 317
AIR SERVICE TESTS 103
AMERICANIZATION IN INDUSTRIES, CONFERENCE ON 269
ANALYSIS OF TRADE ABILITY 352
Applied Eugenics 96
ARMY PERSONNEL WORK 1
ATTITUDES IN DEBATE, MENTAL 156
AVERAGE MENTAL AGE OF ADULTS 317

Child, Kingdom of the 100
CHILDREN IN GOOD AND POOR FARMING DISTRICTS COMPARED, 283; COMPARISON OF COLORED AND WHITE, 277; THEIR INTELLIGENCE RELATED TO FATHER'S OCCUPATION 366
Choosing Employees by Mental and Physical Tests 388
CLASSIFICATION OF MEN, PRINCIPLES UNDERLYING THE 50
CLERKS (OFFICE), STANDARDIZED TEST FOR 248
CLINICAL PSYCHOLOGY, SYMPOSIUM ON 81
COACHPROOF TEST 172
Commercial Tests and How to Use Them 194, 387
COMPARATIVE VARIABILITY OF BOYS AND GIRLS 151
COMPARISON OF: COLORED AND WHITE CHILDREN, 277; CHILDREN IN GOOD AND POOR FARMING DISTRICTS, 283; TWO INTELLIGENCE SCALES, 310
CONFERENCE ON AMERICANIZATION IN INDUSTRIES 269
Creative Impulse in Industry 98
CROSS-OUT TESTS 138
CUED COMBINATION METHOD 183

Deficiency and Delinquency, an Interpretation of Mental Testing 386
DELINQUENTS AND ILLITERATES, INFORMATION TEST FOR, 374; AND DEPENDENTS, MENTAL RATING OF JUVENILE 291

EDUCATIONAL DIFFERENCES AMONG TRADESMEN 33
EFFICIENCY OF A GROUP SCALE OF INTELLIGENCE, 68; FOR SCHOLASTIC PROGNOSIS 381
Eugenics, Applied 96

FILING TEST 178
FLYING, TESTS OF APTITUDE FOR 103

GEOGRAPHICAL INFORMATION IN STUDENTS 186
GROUP EXAMINATION OF INTELLIGENCE INDEPENDENT OF LANGUAGE, 13 and 199; INTELLIGENCE SCALE FOR PRIMARY GRADES, 215; SCALE FOR SCHOLASTIC PROGNOSIS, 381; SCALE OF INTELLIGENCE, ITS PRACTICAL EFFICIENCY 68

INDIVIDUAL RECORDS COMPARED WITH STANDARD FILE 191
INDUSTRIES, CONFERENCE ON AMERICANIZATION IN 269
Industry, Creative Impulse in 98
INFORMATION TEST FOR DELINQUENTS AND ILLITERATES 374
INTELLIGENCE, EFFICIENCY OF A GROUP SCALE OF, 68; GROUP EXAMINATION OF, 13; OF SCHOOL CHILDREN RELATED TO FATHER'S OCCUPATION 366

Intelligence Scale, Absolute, 118; for Primary Grades, 215; Two Scales Compared 310
Intelligence Test, A Non-Language Group 199
Intellectual Development Despite Enforced Seclusion 167

Juvenile Dependents and Delinquents, Rating of 291

Kingdom of the Child 100

Learning Curve in Typewriting 252

Manual for Otis' Group Intelligence Scale 194
Mental Age of Adults 317

Nerve Control and How to Gain It 287
Notes on Bureau of Educational Research, Univ. of Ill., 196; Bureau of Salesmanship, Carnegie Inst. Tech., 197, 290 and 394; Children's Service Bureau, Youngstown, O., 289; Institute for Public Service, N. Y., 290; Iowa Child Welfare Research Station, 394; Mass. Society for Mental Hygiene, 396; National Research Council, 393; Society for American Fellowships in French Universities, 395; Training Course at the Athenaeum Press 396

Omnibus Test for Reasoning 189

Principles Underlying the Classification of Men 50
Psychology and Pedagogy of Anger 287

Rating of Juvenile Dependents and Delinquents 291
Religion and Culture, A Critical Survey 389

Scales of Intelligence, Comparison between Two, 310; Group Intelligence for Primary Grades 215
Standardized Test for Office Clerks 248
Symposium on Clinical Psychology 81

Telegraphers, Mental Tests for 110
Test, A Non-Language Group Intelligence, 199; for Mental Accuracy, 179; for Office Clerks, Standardized 248
Tests for Prospective Telegraphers, 110; of Aptitude for Flying 103
Trade Ability, Analysis of, 352; Test, A Written 358
Tradesmen, Educational Differences among 33
Typewriting, Learning Curve in 252

Variability of Boys and Girls 151

Winston Simplified Dictionary 194
Written Trade Test 358

INDEX OF AUTHORS

(The names of authors of original contributions are printed in SMALL CAPITALS.)

ARTHUR, GRACE 118

BILLS, M. A. 388
BINGHAM, W. V. 1
BLANCHARD, PHYLLIS 100, 287
Bruce, H. Addington 287

CARNEY, CHESTER S. 269
CHAPMAN, J. CROSBY 252, 358
Cody, Sherwin 194, 387

DOLL, EDGAR A. 317

FIELD, I. A. 287
FOSTER, JOSEPHINE CURTIS 167
FRASIER, GEORGE W. 151

GEISSLER, L. R. 98
GESELL, ARNOLD 81
GODDARD, HENRY H. 84
GOULD, MIRIAM C. 96

HAINES, THOMAS H. 291
Hall, G. Stanley 100
Heniger, Alice Minnie Herts 100
HENMON, V. A. C. 103

Johnson, Roswell H. 96

KARLSON, KARL J. 389
KELLEY, C. M. 172
KELLEY, TRUMAN L. 50
Kemble, William Fretz 383
KENAGY, H. C. 387
KITSON, HARRY DEXTER 310

Lewis, William D. 194
LOWELL, FRANCES 215

Marot, Helen 98
Miner, James Burt 386

Otis, Arthur S. 194

PARTLOW, W. D. 291
PAYNTER, RICHARD H., JR. 329
PINTNER, RUDOLF 33, 199
Popenoe, Paul 96
PRESSEY, L. W. 68, 138
PRESSEY, S. L., 68, 138, 277, 283, 366, 374, 381

RALSTON, RUTH 366
Richardson, Roy Franklin 287
ROBINSON, EDWARD S. 352

Schleiter, Frederick 389
SHIVELY, I. M. 374
Singer, Edgar A. 194
SPROWLS, J. W. 386
SWIFT, WALTER B. 156

TETER, G. F. 277
THOMAS, J. B. 283
THORNDIKE, EDWARD L. 13
THURSTONE, L. L. 110, 248
TOOPS, HERBERT A. 33, 358

WALLIN, J. E. WALLACE 87
WELLS, F. L. 172
WOODROW, HERBERT 118

Directory of American Psychological Periodicals

American Journal of Psychology—Worcester, Mass.: Florence Chandler.
Subscription $5. 600 pages annually. Edited by G. Stanley Hall.
Quarterly. General and experimental psychology. Founded 1887.

Pedagogical Seminary—Worcester, Mass.: Florence Chandler.
Subscription $5. 575 pages annually. Edited by G. Stanley Hall.
Quarterly. Pedagogy and educational psychology. Founded 1891.

Psychological Review—Princeton, N. J.: Psychological Review Company.
Subscription (with Psychological Bulletin) $6. 480 pages annually.
Bi-monthly. General. Founded 1894. Edited by Howard C. Warren.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company.
Subscription $3.25. 480 pages annually. Psychological literature.
Monthly. Founded 1904. Edited by Shepherd I. Franz.

Psychological Monographs—Princeton, N. J.: Psychological Review Company.
Subscription $4. per vol. 500 pages. Founded 1895. Edited by James R. Angell.
Published without fixed dates, each issue one or more researches.

Psychological Index—Princeton, N. J.: Psychological Review Company.
Subscription $1.25 200 pages. Founded 1895. Edited by Madison Bentley.
An annual bibliography of psychological literature.

Journal of Philosophy, Psychology and Scientific Methods—New York:
Sub-Station 84. Bi-weekly. 728 pages per volume. Founded 1904.
Subscription $4. Edited by F. J. E. Woodbridge and Wendell T. Bush.

Archives of Psychology—Sub-Station 84, New York: Archives of Psychology.
Subscription $5. 600 pp. annually. Founded 1906. Edited by R. S. Woodworth.
Published without fixed dates, each number a single experimental study.

Journal of Abnormal Psychology—Boston: Richard G. Badger. Subscription $5. 480 pages annually. Ed. by Morton Prince. Bi-monthly. Founded 1906.
Entire field of abnormal psychology.

Psychological Clinic—Philadelphia: Psychological Clinic Press.
Subscription $1.50. 280 pages annually. Edited by Lightner Witmer.
Monthly (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Training School Bulletin—Vineland, N. J.: The Training School.
Subscription $1. 160 pages annually. Ed. by E. R. Johnstone. Founded 1904.
Monthly (10 numbers). Psychology and training of defectives.

Journal of Race Development—Worcester, Mass.: Louis N. Wilson.
Subscription $3. 460 pages annually. Founded 1910. Quarterly.
Edited by George H. Blakeslee and G. Stanley Hall.

Journal of Educational Psychology—Baltimore: Warwick & York.
Subscription $3. 600 pages annually. Founded 1910.
Monthly (10 numbers). Managing Editor, J. Carleton Bell.
(Educational Psychology Monographs. Edited by Guy M. Whipple.
Published separately at varying prices. Same publishers.)

Journal of Animal Behavior—Cambridge, Mass.: Emerson Hall.
Subscription $5. 450 pages annually. Founded 1911.
Bi-monthly. Robert M. Yerkes, Managing Editor.

The Behavior Monographs—Cambridge, Mass.: Emerson Hall.
Subscription $5. 450 pages per volume. Edited by John B. Watson.
Published without fixed dates, each number a single research.

Psychoanalytic Review—Washington, D. C.: 3617 Tenth St., N. W.
Subscription $5. 500 pages annually. Psychoanalysis.
Quarterly. Founded 1913. Ed. by W. A. White and S. E. Jelliffe.

Journal of Experimental Psychology—Princeton, N. J.: Psychological Review Company. 480 pages annually. Experimental. Founded 1916.
Subscription $3.25 Bi-monthly. Edited by John B. Watson.

Journal of Applied Psychology—Worcester, Mass.: Florence Chandler.
Subscription $4. 400 pages annually. Quarterly. Founded 1917.
Edited by G. Stanley Hall and L. R. Geissler.

THE BRANDOW PRINTING COMPANY,
ALBANY, N. Y.